Nonprogressive Diffusion on Social Networks: Approximation and Applications

Yunduan Lin
CUHK Business School, The Chinese University of Hong Kong, Hong Kong, China, yunduanlin@cuhk.edu.hk

Heng Zhang
W. P. Carey School of Business, Arizona State University, hengzhang24@asu.edu

Renyu Zhang
CUHK Business School, The Chinese University of Hong Kong, Hong Kong, China, philipzhang@cuhk.edu.hk

Zuo-Jun Max Shen
Faculty of Engineering and Faculty of Business and Economics, University of Hong Kong, Hong Kong, China, maxshen@hku.hk

Nonprogressive diffusion models the spread of behavior on social networks, where agents are allowed to reverse their decisions as time evolves. To provide an efficient framework for evaluating and optimizing nonprogressive diffusion, we introduce a comprehensive model along with a Fixed-Point Approximation (FPA) scheme, which admits both theoretical guarantee and computational efficiency. We show that the approximation error depends on the network structure and derive order-optimal bounds for this error based on a newly proposed network measure. Additionally, we propose two easy-to-calculate network metrics, one at the node level and the other at the network level, that serve as reliable indicators of FPA performance. Our results indicate that the FPA scheme is particularly accurate for dense and large networks, which are typically challenging to analyze via simulation. To showcase the broad applicability of our approach, we apply the FPA scheme to well-known problems like influence maximization and optimal pricing on social networks. Finally, we conduct extensive numerical experiments on both synthetic and real-world networks. On real-world networks, the FPA scheme achieves computational speedups of 70-230 times compared to naïve agent-based simulation and 23-30 times compared to more advanced simulation method, while maintaining a mean absolute percentage error of less than 3.48%.

Key words: Nonprogressive network diffusion, Large-scale network approximation, Network centrality, Influence maximization, Pricing

1. Introduction

Social networks fundamentally shape our lives. People are more receptive to information shared by their friends and relatives (Lu et al. 2013) and more inclined to make a purchase when informed by their acquaintances (Ma et al. 2015, Bapna and Umyarov 2015). It is even more so in the digital era—globally, 4.62 billion people, approximately 58.4% of the population worldwide, used online social network platforms, such as Facebook, YouTube, and Tiktok, by January 2022 (Datareportal

2022). These platforms extend the reach and complexity of our social networks, as both friends and strangers online contribute to shaping our opinions and choices.

Within such networks, each agent both affects and is affected by others, setting the stage for the diffusion of information and behavior. In this work, we use the term diffusion to refer to the spread of information or behavior when agents are influenced by their social connections. Platforms that leverage network diffusion can substantially boost their impact and profitability (Shriver et al. 2013). However, understanding network diffusion is a complex undertaking. It involves not just individual preferences, but also the intricate relationships that bind them. This complexity has made network analysis an enduring area of study, engaging generations of researchers (see books Jackson 2010).

Network diffusion analysis spans multiple domains, such as computer science (Kempe et al. 2003, Acemoğlu et al. 2013), economics (Sadler 2020, Acemoglu et al. 2011), operations management (Song and Zipkin 2009, Candogan et al. 2012, Shen et al. 2017, Wang and Wang 2017) and epidemiology (Kermack and McKendrick 1927, Drakopoulos and Zheng 2017). In the seminal paper (Kempe et al. 2003), diffusion processes are broadly categorized into progressive and nonprogressive types. While progressive diffusion deals with unidirectional changes in the state, such as adopting new technology or purchasing a product, our study focuses on *nonprogressive diffusion*, which allows for bidirectional state transitions. This type is particularly relevant for scenarios when decisions can be reversed, such as subscription to a membership, belief propagation, or infection dynamics in a pandemic.

Analyzing network diffusion primarily adheres to one of two approaches. The first is microfounded by capturing the concrete network topology and the stochastic evolution of agent states. Notable models of this include the independent cascade model (Goldenberg et al. 2001) and the linear threshold model (Granovetter 1978, Schelling 1978). While these models capture fine granularity, detailing the diffusion on an individual basis over time, their intricate nature often leads to computational challenges. In most cases, simulation happens to be the only viable tool to analyze such models, making the optimization, even for a sparse and moderate-sized network, time-consuming (Chen et al. 2009). Conversely, the second approach offers a macroscopic view, simplifying the diffusion process. Some models (e.g., Bass models; Bass 1969) bypass the intricacies of network topology, focusing on the overall population. Others (e.g., Candogan et al. 2012, Jackson et al. 2020) ignore the stochasticity and focus on the equilibrium outcome. This macro lens, while sacrificing detailed characterization of the diffusion, facilitates efficient analyses and generates sharper insights.

Our work bridges these two approaches in the context of nonprogressive diffusion by providing a simple, efficient, and accurate approximation scheme. We base on a general microfounded diffusion model that considers heterogeneous agents, local network effects, and network topology. While analytically characterizing the adoption rate over time for each agent in such a detailed model is technically intractable, we propose a fixed-point approximation (FPA) scheme that provides an

effective approximation through a set of easily solvable fixed-point equations. Notably, we show that the FPA scheme comes with provable guarantees. Its performance is associated with the network structure and improves for larger and denser networks. We also propose easy-to-calculate metrics at both node and network levels that can efficiently indicate FPA performance for different network structures. Moreover, the FPA scheme paves the way for optimizing operational decisions, such as the influence maximization and optimal (dynamic) pricing problems in the nonprogressive diffusion context. It enables straightforward problem formulation and algorithm development, that are not just computationally efficient but also yield near-optimal solutions. In summary, through the FPA scheme, we show that the diffusion characterized by a "micro-model" can be accurately approximated by an easy-to-analyze "macro-model", integrating the advantages of both modeling paradigms.

1.1. Contributions and Organization

Our research contributions are summarized as follows:

- **Provable approximation scheme for a general nonprogressive diffusion model.** We investigate nonprogressive diffusion through a micro-founded, dynamic and stochastic model. Our model generalizes the well-known linear threshold (LT) model and adapts it to the nonprogressive diffusion. Based on this model, we propose the FPA scheme to approximate the adoption probability of each agent. To validate this approach, we develop a nontrivial "fixed-point sandwich" technique, establishing an order-optimal error bound of the FPA scheme in terms of a newly proposed network measure. The bound indicates its superior performance for large and dense networks, which are otherwise challenging to simulate. We further connect the error bound to two novel and easy-to-calculate metrics: the *inverse in-degree centrality* and the *inverse in-degree density*. These metrics, which we propose to gauge FPA performance, provide valuable insights into both node-level and network-wide structures. They serve as reliable indicators for FPA performance in diverse network configurations. Our extensive empirical studies further confirm the theoretical results. For real-world networks, it achieves a mean absolute percentage error of less than 3.48% among all instances tested, while concurrently accelerating computation by factors ranging from 70-230, compared to a conventional approach such as the naïve agent-based simulation.
- **Wide applicability in optimizing operational decisions.** The FPA scheme offers a powerful tool to reformulate and solve operational decision-making problems in the nonprogressive diffusion setting. Leveraging our approximation error bound, the reformulated problems lead to efficient algorithms that guarantee high-quality decisions, providing a clear advantage over simulation-based methods. We illustrate this with two examples: influence maximization (IM) and optimal (dynamic) pricing. For the IM problem, we show that, under technical conditions,

the influence function is submodular with regard to the seed set in the reformulated problem. This extends the greedy algorithm to more general settings, significantly improving efficiency. For the optimal pricing problem, the FPA scheme facilitates near-optimal solutions for general static pricing problems, which accounts for the limiting agent behavior. Specifically, we derive approximate gradients that enable the efficient computation of near-optimal solutions. When perfect price discrimination is implemented, we can reformulate the static pricing problem as a convex optimization problem under specific conditions, yielding the optimal solution. Additionally, the dynamic pricing problem, considering transient agent behavior, can also achieve the optimal solution within the same framework under perfect price discrimination.

The remainder of this paper is structured as follows: In this section, we continue to review related literature. Section 2 introduces the nonprogressive diffusion model and characterizes the adoption probabilities. In Section 3, we describe the FPA scheme and present our main theoretical results. In Section 4, we establish the order-optimal error bound, followed by extensive numerical experiments in Section 5. We apply the FPA scheme to the IM and pricing problems in Section 6. Section 7 concludes this paper. Throughout, we use increasing and decreasing in the non-strict sense.

1.2. Literature Review

Our paper contributes to network diffusion literature, focusing on diffusion models and related optimization problems. Below, we provide a review to delineate how our work relates to and diverges from existing studies.

Diffusion Models. Various models have been proposed across disciplines to characterize diffusion for specific applications. However, a consistent trade-off can be observed: researchers often have to choose between a comprehensive model and practical efficiency. For instance, the LT model (Granovetter 1978, Schelling 1978) incorporates the network structure but is computationally challenging, as evidenced by Chen et al. (2010). In contrast, the Bass model (Bass 1969) abstracts away most information about the network structure and individual agents, but benefits from analytical tractability, offering closed-form expressions for critical values that facilitate optimization (Agrawal et al. 2021, Lin et al. 2021). Our review categorizes the existing literature into two groups: models that are inherently complex and computationally burdensome, which align with our micro-founded model, and models that prioritize analytical tractability, which align with our FPA scheme. Additionally, we discuss how traditional approaches, such as mean-field approximations, have been used to mitigate the computational challenges associated with the first group of models.

Our model builds upon the LT model and is particularly aligned with the nonprogressive LT variant[1] introduced by Kempe et al. (2003). This variant preserves many features of the progressive

[1] Hereafter, we will refer to this as the nonprogressive LT model to differentiate it from its progressive counterpart.

LT model but introduces flexibility by allowing random thresholds to vary independently at each time step, in contrast to the fixed random thresholds in the progressive version. We extend the nonprogressive LT model further by introducing agent heterogeneity and accommodating a broader class of randomness distributions, thereby enhancing its applicability to diverse diffusion contexts.

Several parallel streams of work share conceptual connections with our micro-founded model. In the field of social learning, agents form beliefs about a binary signal based on their neighbors' beliefs and often converge toward an equilibrium (Jadbabaie et al. 2012, Chandrasekhar et al. 2020, Allon et al. 2019). While these models emphasize the learning process and the final belief distribution, our work characterizes individual adoptions in arbitrary network structures to facilitate operational decision-making. Moreover, while our framework can be conceptualized as a Markov chain on graphs, it fundamentally differs from random walks on the graph (Göbel and Jagers 1974), which track the movement of a single agent across graph vertices. Our framework focuses on the diffusion process across the network, requiring an analysis of how the agent states evolve collectively over time. Models like Markov Random Fields also share similarities, as they define conditional dependencies among variables on a graph. For example, the Ising model (Ising 1924) in statistical mechanics describes binary states of vertices influenced by their neighbors, typically aiming to estimate marginal distributions at equilibrium. Unlike these models, our framework formulates a Markov chain to model the dynamic diffusion process, providing a micro-foundation for understanding individual agent behaviors by incorporating temporal evolution alongside spatial dependencies.

Another stream of models emphasizes analytical traceability, often at the expense of descriptive power, aligning conceptually with our FPA scheme. For instance, many engineering and economics applications describe the interactions using network games (e.g., see Ballester et al. 2006, Candogan et al. 2012, Afèche et al. 2023, Baron et al. 2022, Feng et al. 2022). A central goal of this literature is to analyze various equilibria. Although our fixed-point approximation is reminiscent of the equilibrium in the network games, our focus diverges in its relation to a concrete micro-founded model. Similarly, a number of operations management studies incorporate network externality into consumer choice models. This type of work, serving for the subsequent assortment or pricing problem, often simplifies the model. For example, some consider only global effects by looking at market-wide adoption averages (Du et al. 2016, Wang and Wang 2017), while others use myopic local proxy or specific types of networks (Gopalakrishnan et al. 2022, Xie and Wang 2020). In contrast, our framework considers full network information and operates in a more general setting.

Finally, our work contributes to the literature on approximation schemes for network diffusion models. While conceptually related to mean-field approximations (Benaïm and Weibull 2003, Van Mieghem et al. 2008), which typically offer deterministic descriptions at the population level and focus on equilibrium outcomes, our approach goes further by approximating the entire dynamic

diffusion process. In Markov random field models, where the equilibrium is well-defined, mean-field approximations (Boykov et al. 1998, Yeomans 1992) reduce complexity by fitting deterministic approximations into a predefined equilibrium representation. However, our model faces the challenge of evolving correlations among agents behavior over time. This requires a fundamentally different approach to approximation and analysis, with tools that are distinct from traditional methods. This dynamic consideration not only introduces technical complexities but also enables us to bridge the aforementioned two distinct streams of models. Our FPA scheme facilitates computationally efficient analysis while preserving the key features of the underlying micro-founded model, providing a dual perspective for studying and understanding network diffusion processes.

Optimization with Network Diffusion. The FPA scheme has a wide range of applications. In this paper, we highlight its use in two examples, influence maximization and optimal pricing problems, and provide a concise review of the literature on these topics. Kempe et al. (2003) first consider the issue of choosing an influential set of seed agents to maximize total influence as a discrete optimization problem. They show that the IM problem, under the LT model, is NP-hard for both progressive and nonprogressive cases. Moreover, evaluating the total influence for different seed sets requires extensive simulations, making it time-consuming to obtain even approximate solutions. We refer readers to the survey (Li et al. 2018) for a comprehensive review of existing approaches. These approaches compromise either accuracy or efficiency and are not ideal for practical use. With the FPA scheme, we can effectively balance both. For the pricing problem, there is a growing literature in the economics and operations management communities that considers the presence of network effects (Anari et al. 2010, Hu et al. 2020, Li 2020, Yang and Zhang 2022, Huang et al. 2022). Recent studies on the single-item pricing problem with the network effect can be found in Candogan et al. (2012), Du et al. (2018), and Nosrat et al. (2021). Compared with others, our framework and the proposed algorithms can be used to consider a more general setting. Moreover, much of this literature focuses on the static pricing problem, aiming to determine an equilibrium price while considering network effects. In this work, we also address the dynamic pricing problem under certain conditions, which allows us to additionally incorporate transient behaviors before reaching equilibrium.

2. Nonprogressive Network Diffusion Model

In this section, we first introduce the nonprogressive diffusion model and then characterize the limiting behavior of each agent. While this model can be applied to various nonprogressive diffusion settings, we use service adoption on an online social network platform for illustration.

2.1. Preliminaries and Formulation

We model the social network platform (e.g., TikTok) as a graph $G = (V, E)$ with n agents, where $V := \{1, 2, ..., n\}$ is the set of agents and $E := \{1, 2..., |E|\}$ is the set of directed edges. A directed

edge $(i,j) \in E$, where $i,j \in V$, implies that agent j is influenced by agent i, and we call i an *in-neighbor* of j. We interpret $(i,j) \in E$ as j following i on the platform. We use $\mathcal{N}_i$ to denote the set of all in-neighbors for agent i (i.e., $\mathcal{N}_i := \{j \in V : (j,i) \in E\}$) and $d_i := |\mathcal{N}_i|$ to denote the in-degree (i.e., the number of in-neighbors). Throughout, we use agent and node interchangeably.

We use t to denote the discrete time period, starting with $t = 0$ as the service launch time. Define $Y_i(t) \in \{0,1\}$ as the state of agent i at time t, where $Y_i(t) = 1$ (resp. $Y_i(t) = 0$) means adoption (resp. nonadoption) of the service in this period. The initial state $\mathbf{Y}(0)$ follows an arbitrary distribution on $\{0,1\}^n$. For all $t \geq 1$, each agent i decides whether to adopt the service based on their realized utility $u_i(t)$ during that period, which is given by

$$u_i(t) := v_i + \beta \cdot \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} + \epsilon_i(t). \tag{1}$$

Without loss of generality, we normalize the utility of nonadoption to 0, so that $Y_i(t) = \mathbb{1}\{u_i(t) \geq 0\}$. As shown in (1), $u_i(t)$ consists of three parts: the idiosyncratic intrinsic value v_i, *the local network effect* $\beta \cdot \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i}$, and random noise $\epsilon_i(t)$. The value v_i reveals the personalized preference and remains constant over time. Analytically, it can be estimated from features such as demographics and behavioral data with the support of big data. It may also be affected by the platform strategies. For example, the price of a paid service (e.g., YouTube Premium) will definitely affect whether and how, the agent likes it. The local network effect term captures peer influence, with β quantifying the network effect intensity. If agent i has no in-neighbors (i.e., $\mathcal{N}_i = \emptyset$), we set this term to 0. Finally, we assume the random noise $\epsilon_i(t)$ is independent and identically distributed (i.i.d.) across agents and time, with $\mathbb{E}[\epsilon_i(t)] = 0$. While the i.i.d. assumption is standard, we acknowledge that real-world scenarios may introduce more complexity. For example, agents may gradually reveal their preferences, leading to a diminishing variance in random noise over time. Alternatively, public news may introduce positively correlated noise across agents at the same time. Despite these complexities, the approximation approach proposed later in this work remains valid, though the approximation error may increase. Appendix C provides a detailed discussion on relaxing the i.i.d. assumption. For now, we impose no further constraints on its distribution, except for the following mild condition.

ASSUMPTION 1 **(Lipschitz Continuity)**. *The random noise $\epsilon_i(t)$ has an L-Lipschitz continuous cumulative distribution function (CDF): $|F_\epsilon(x) - F_\epsilon(y)| \leq L|x-y|$ for any $x, y \in \mathbb{R}$.*

This assumption ensures a sufficiently smooth noise distribution. Assumption 1 holds for any continuous distribution with a bounded probability density function (PDF), such as uniform, logistic, or normal distribution, making them compatible with our model. The parameter L characterizes the sensitivity of agent behavior to random noise. We note that our result remains valid even if L is not uniformly defined over the entire CDF, as assumed in Assumption 1. Specifically, for a given

intrinsic value $\mathbf{v}$, one can define an instance-dependent L that need not be the Lipschitz constant over the entire real line but only within the domain $[-\max_{i\in V} v_i - \beta, -\min_{i\in V} v_i]$. In this case, our results still hold. However, for clarity and consistency, we maintain Assumption 1 throughout.

A natural goal of this model is to quantify the total diffusion in the network. In line with prior studies (Kempe et al. 2003), we focus on the limiting adoption probability[2]. Provided it converges, this also represents the accumulated reward (frequency of adoptions) in the long run. However, to meaningfully discuss such behavior, we must first ensure its existence. Without additional assumptions, the process may diverge or yield multiple outcomes, complicating the focus on a single outcome of interest. As noted by Lee et al. (2014) in similar contexts, even identifying the number of equilibria in general network structures can be challenging. To guarantee a unique outcome, we impose Assumption 2, which excludes divergent or periodic behavior (Proposition 1) and ensures the existence of a valid fixed-point solution for our approximation (Proposition 2).

ASSUMPTION 2 **(Bounded Network Effects)**. *The network effect intensity satisfies* $|\beta| < 1/L$.

This assumption ensures that agents' behavior is not overly sensitive to network effects or random noise. It is well-established in the literature and is typically satisfied in practice. Specifically, it aligns with conditions in the network diffusion literature (Kempe et al. 2003) and assumptions in social economics research (e.g., see Horst and Scheinkman 2006, Wang and Wang 2017, Xu 2018, Jackson et al. 2020, Gopalakrishnan et al. 2022). Empirical studies that calibrate β in real-world applications further support the validity of this assumption. For instance, Wang and Wang (2017) assumes random noise follows a standard logistic distribution Logistic$(0,1)$[3] (logit model), resulting in $L = 1/4$, with calibrated β ranging from 1.391 to 2.392. Similarly, Chen et al. (2021) assumes a standard normal distribution (probit model), yielding $L = 1/\sqrt{2\pi}$, with calibrated β ranging from 0.69 to 1.11. These empirical results provide strong evidence that the assumption holds in practice.

While our theoretical analysis is based on Assumption 2, we acknowledge that it may not always hold. Therefore, Section 5.1 extends our discussion with numerical experiments that explore the implications of violating Assumption 2. For the rest of our theoretical analysis, we assume positive network effects, i.e., $0 < \beta < \frac{1}{L}$. However, our results can also be generalized to account for negative network effects, where $-\frac{1}{L} < \beta < 0$.

We remark on the notations. Hereafter, we use a bold math notation to denote the collection of a particular variable over all agents in vector form. We represent a specific diffusion instance by a quadruple $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, since the network structure and intrinsic values identify a diffusion case, meanwhile, the noise distribution and the network effect intensity constitute the diffusion environment. Sequences of such instances are represented by a series of these quadruples.

[2] While our primary results emphasize the limiting behavior, they can also be extended to capture the entire diffusion process, including transient behaviors.

[3] The CDF of Logistic$(\mu;\sigma)$ is $F_\epsilon(x) = 1/(1+\exp\{-(x-\mu)/\sigma\})$.

2.2. A Markov Chain Perspective

Notably, each diffusion instance can be characterized by a Markov chain (MC), of which the state space is the set of indicator vectors denoting all possible combinations of adoption decisions, represented by $\{0,1\}^n$. The transition probability from state $\mathbf{y}$ to $\mathbf{y}'$ can be computed as

$$P(\mathbf{y},\mathbf{y}') \;=\; \prod_{i\in V} \mathbb{P}(Y_i(t)=y_i' | \mathbf{Y}(t-1)=\mathbf{y}) \;=\; \prod_{i\in V} F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{d_i}\right)^{1-y_i'} \cdot \left[1-F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{d_i}\right)\right]^{y_i'}.$$

Our primary interest lies not in the individual MC states, but rather in the overall adoption probability for each agent. To that end, we define the adoption probability of agent i at time t as

$$q_i(t) \;:=\; \mathbb{P}(Y_i(t)=1) \;=\; \sum_{\mathbf{y}\in\{0,1\}^n} \mathbb{1}\{y_i=1\}\cdot\mathbb{P}(\mathbf{Y}(t)=\mathbf{y}). \tag{2}$$

We have the following proposition on the limiting behavior of $\mathbf{q}(t)$ when t tends to infinity.

PROPOSITION 1 **(Limiting Adoption Probability)**. *Under Assumptions 1 and 2, for any initial state $\mathbf{Y}(0)\in\{0,1\}^n$, the adoption probability of each agent i converges to*

$$\lim_{t\to\infty} q_i(t) = q_i^* := \sum_{\mathbf{y}\in\{0,1\}^n} \mathbb{1}\{y_i=1\}\cdot\pi(\mathbf{y}),$$

as t tends to infinity, where $\boldsymbol{\pi}$ is the stationary distribution of the MC that satisfies $\boldsymbol{\pi}=\boldsymbol{\pi}P$.

As shown in the proof of Proposition 1, this MC only has a *single aperiodic recurrent class*. Thus, limiting distribution $\boldsymbol{\pi}$ leads to the existence of $\mathbf{q}^*$. By the standard MC theory, one can verify that

$$\lim_{t\to\infty}\frac{1}{t}\cdot\sum_{\tau=1}^{t} Y_i(\tau) = q_i^* \quad \text{a.s.} \quad \text{and} \quad \lim_{t\to\infty}\frac{1}{t}\cdot\sum_{\tau=1}^{t} q_i(\tau) = q_i^*, \quad \forall i\in V, \tag{3}$$

for any initial state $\mathbf{Y}(0)$. As a result, various operational problems, such as the IM (Section 6.1) and optimal pricing problem (Section 6.2), can be generally framed as:

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \qquad g\Big(\mathbf{q}^*(G,\mathbf{v}(\mathbf{x}),F_\epsilon(\cdot),\beta),\mathbf{x}\Big). \tag{4}$$

Here, $\mathbf{x}$ represents platform decisions, with $\mathcal{X}$ denoting its feasible set. For simplicity, we only consider decisions that influence diffusion outcomes by altering intrinsic values. With a slight abuse of notation, $\mathbf{v}(\cdot)$ represents intrinsic values as a function of platform decisions, and $\mathbf{q}^*(\cdot)$ denotes the mapping from a diffusion instance to the limiting probability vector. Finally, $g(\cdot,\cdot)$ is the objective function that depends on the diffusion outcome $\mathbf{q}^*$ and decision $\mathbf{x}$. For example, the IM problem can be formulated as (4), where the decision $\mathbf{x}$ is setting the intrinsic utility of a set of seed users to sufficiently high levels, and the objective $g(\cdot,\cdot)$ is the expected total limiting adoptions, $\sum_{i\in V} q_i^*$. For the optimal pricing problem, the decision $\mathbf{x}$ is the price vector that affects the intrinsic value of

each agent, and the objective $g(\cdot,\cdot)$ is the expected profit under the limiting adoption probability, i.e., $\sum_{i\in V} q_i^* x_i$. The specific formulations of these problems will be presented in Section 6.

Solving problem (4) is challenging, due to the absence of closed-form expressions for $\mathbf{q}^*$ and the exponential growth of the MC states. It is intractable to construct the transition matrix even for a moderate-sized network, let alone to calculate $\mathbf{q}^*$. Therefore, problem (4) is generally intractable analytically, which motivates us to develop our approximation scheme presented in Section 3.

3. Fixed-Point Approximation (FPA) Scheme

In this section, we introduce the FPA scheme and provide a comprehensive analysis of its performance, including upper bounds for approximation errors. Additionally, we also present two easy-to-compute metrics designed to evaluate FPA performance; see (6).

3.1. Overview and Motivating Example

For a given diffusion instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, we will show that the limiting adoption probability $\mathbf{q}^*$ can be reasonably approximated by the solution $\boldsymbol{\mu}^*$ of the following system of equations:

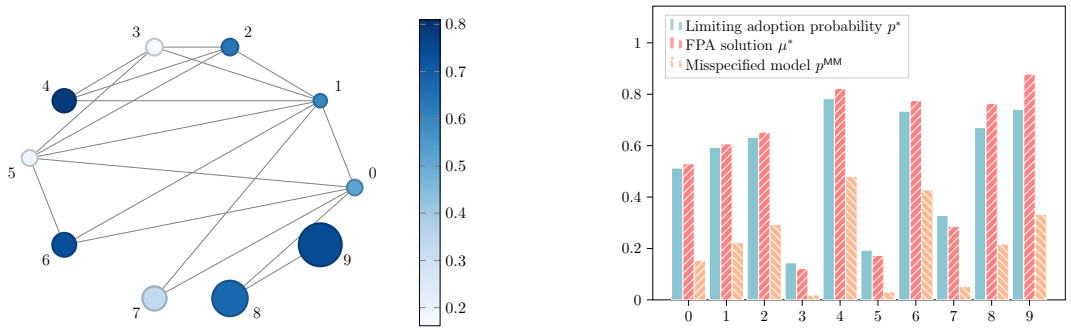
$$\mu_i = 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \mu_j}{d_i}\right), \text{ for all } i \in V. \tag{5}$$

While our main result focuses on approximating the limiting adoption probability, the full approximation framework and theoretical analysis go beyond this. Specifically, it captures the entire diffusion process, including transient behaviors. Since (5) represents the limiting state of a fixed-point iteration, which corresponds to a deterministic dynamical system, the FPA scheme allows us to track the evolution of this system and approximate the entire diffusion process. As such, the FPA scheme is broadly applicable to network diffusion problems, including those with transient behaviors.

We begin with a motivating example to illustrate the quality of FPA, as shown in Figure 1. The example highlights heterogeneity in both network connectivity and intrinsic values among agents. For detailed information on this example instance, including numerical results, please refer to Appendix E.1. In Figure 1, to provide an intuitive understanding of the network effects, we also introduce a misspecified model (MM) as a benchmark. For the MM, the adoption probability for agent i is calculated as $q_i^{\mathrm{MM}} := \mathbb{E}[\mathbb{1}\{v_i + \epsilon_i \geq 0\}]$, ignoring the network effects.

Figure 1a presents the network structure and approximation results for the example. Clearly, nodes with fewer neighbors exhibit larger errors, while well-connected nodes yield smaller errors. Figure 1b compares the values of $\mathbf{q}^*$, $\boldsymbol{\mu}^*$ and $\mathbf{q}^{\mathsf{MM}}$, showing the impact of network effects, as evidenced by the discrepancy between $\mathbf{q}^*$ and $\mathbf{q}^{\mathsf{MM}}$. Using $\mathbf{q}^*$ as the baseline, the mean absolute errors for $\boldsymbol{\mu}^*$ and $\mathbf{q}^{\mathsf{MM}}$ are 0.045 and 0.310, respectively. These findings confirm the feasibility of the FPA scheme and suggest that the approximation is more accurate for agents with central positions in the network.

Next, we theoretically analyze the deviation between $\mathbf{q}^*$ and $\boldsymbol{\mu}^*$, focusing on how it depends on network structure. The key challenge lies in the temporal and spatial correlations in adoptions, which are further complicated by the nonlinear evolution introduced by the general noise distribution.



(a) Approximation results on 10-node example instance

(b) Values of adoption probabilities

Figure 1 **A 10-node example to illustrate the FPA scheme. In the left subfigure, for each agent i, the color denotes the true values of q_i^* and the size denotes the absolute error $|q_i^* - \mu_i^*|$.**

3.2. The Approximation Error

We first remark on the notation before formally analyzing the error bound. Given a network $G = (V, E)$, we define the matrix $\tilde{\mathbf{A}} \in \mathbb{R}^{n\times n}$ such that $\tilde{A}_{ij} = \frac{1}{d_i}$ if an edge is directed from j to i, and $\tilde{A}_{ij} = 0$ otherwise. This matrix is a scaled version of the adjacency matrix $\mathbf{A}$, where $A_{ij} = 1$ if there is an edge directing from i to j and $A_{ij} = 0$ otherwise. One obtains $\tilde{\mathbf{A}}$ by scaling row i of $\mathbf{A}^\top$ by $\frac{1}{d_i}$. Notably, $\tilde{\mathbf{A}}$ is a row-stochastic matrix, meaning that $\tilde{\mathbf{A}}\mathbf{e} = \mathbf{e}$ where $\mathbf{e}$ is a vector of ones. We also introduce the vector $\mathbf{b} := \left(\frac{1}{d_1}, \frac{1}{d_2}, \ldots, \frac{1}{d_n}\right)^\top$, which contains the reciprocals of each node's in-degree. Lastly, we define $\rho := L\beta$ and let $d_{\min}$ be the minimum in-degree of the network, with $d_{\min} > 0$.

We further introduce two metrics in our analysis, which we term as the *inverse in-degree centrality* $\mathcal{C}(G;\rho)$ and the *inverse in-degree density* $\mathcal{D}(G)$. They are defined as follows:

$$\mathcal{C}(G;\rho) := (1-\rho)\left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell\right)\mathbf{b} = (1-\rho)\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1}\mathbf{b} \quad \text{and} \quad \mathcal{D}(G) := \frac{\mathbf{e}^\top\mathbf{b}}{n}. \tag{6}$$

The *inverse in-degree centrality* $\mathcal{C}(G;\rho)$ is an n-dimensional vector that captures the centrality of each agent, with its i-th entry denoted by $\mathcal{C}_i(G;\rho)$. It bears similarities with the classical Bonacich centrality (Bonacich 1987), taking the form of Neumann series. However, by incorporating the inverse of in-degrees, it is tailored to evaluate the FPA scheme. This centrality is well defined since $\tilde{\mathbf{A}}$ is row-stochastic and $\rho < 1$ by Assumption 2. The *inverse in-degree density* $\mathcal{D}(G)$ is a scalar that represents the average inverse in-degree of all agents, serving as an aggregate measure of FPA performance. Together, these two metrics offer both individualized and holistic views on network structure. This dual perspective not only enhances our understanding of the FPA scheme, but also provides actionable insights into its application across different network configurations.

For clarity, we adopt the subscript $_{\text{ew}}$ to represent entrywise operations on vectors. For instance, for vector $\mathbf{q}$, we define $|\mathbf{q}|_{\text{ew}} := (|q_1|, |q_2|, \ldots, |q_n|)^\top$ and $\mathbf{q}^{\frac{1}{2}}_{\text{ew}} := (\sqrt{q_1}, \sqrt{q_2}, \ldots, \sqrt{q_n})^\top$. Let $\odot$ denote the Hadamard product, which essentially is componentwise product of two matrix. Additionally, we also define the constant $C_\rho := \frac{\rho}{(1-\rho)^{3/2}}$. We now present our key technical result in Theorem 1.

THEOREM 1 **(Entrywise Error Bound of the FPA Scheme)**. *Under Assumptions 1 and 2, for any diffusion instance* $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$*, the absolute difference between the limiting adoption probability* $\mathbf{q}^*$ *and the fixed-point solution* $\boldsymbol{\mu}^*$ *can be upper bounded by*

$$|\mathbf{q}^* - \boldsymbol{\mu}^*|_{\text{ew}} \;\le\; \rho(\mathbf{I} - \rho\tilde{\mathbf{A}})^{-1}\,[\mathcal{B}(G;\rho)]_{\text{ew}}^{\frac{1}{2}} \;\le\; C_\rho \cdot \big[\mathcal{C}(G;\rho)\big]_{\text{ew}}^{\frac{1}{2}},$$

where $\mathcal{B}(G;\rho) := \Big(\sum_{\ell=1}^{\infty} \rho^{2\ell-2}\,\tilde{\mathbf{A}}^{\,\ell} \odot \tilde{\mathbf{A}}^{\,\ell}\Big)\,\mathbf{e}$.

Theorem 1 characterizes the entrywise error between the limiting probability $\mathbf{q}^*$ and our FPA solution $\boldsymbol{\mu}^*$. We show that this error is closely related to the network structure. Specifically, it is linked to the inverse in-degree centrality $\mathcal{C}(G;\rho)$ via an intermediate measure $\mathcal{B}(G;\rho)$. Although $\mathcal{B}(G;\rho)$ involves an infinite sum and cannot be computed efficiently, it offers a detailed characterization of how the error propagates across the network. Moreover, it plays a key role in establishing the order-optimal results discussed later in Section 4.2. In contrast, the inverse in-degree centrality $\mathcal{C}(G;\rho)$ offers a more concise and interpretable metric to quantify the error. Each component of $\mathcal{C}(G;\rho)$ can be viewed as a weighted sum of entries in $\mathbf{b}$, the vector of inverse in-degrees. These weights reflect the connectivity between nodes. Specifically, the weight associated with inverse in-degree b_j in the i-th term of $\mathcal{C}(G;\rho)$ is $(1-\rho)\sum_{P\in\mathcal{P}_{(j,i)}} \rho^{|P|}\prod_{k\in P}\frac{1}{d_k}$, where $\mathcal{P}_{(j,i)}$ denotes the set of directed paths from agent j to i. Remarkably, this weight decays exponentially with the length of the directed path. As a result, the inverse in-degree centrality of each node is predominantly affected by the inverse in-degrees of its nearby nodes. This result rationalizes the observations in Figure 1.

To further understand the intuition of $\mathcal{C}(G;\rho)$, we let $d_{\min,i}(\ell)$ be the minimum in-degree of any node j that is connected to node i via a path of length ℓ. It is straightforward that $d_{\min,i}(\ell) \ge d_{\min}$. By expanding the definition of inverse in-degree centrality, it then holds that

$$\mathcal{C}_i(G;\rho) \le (1-\rho)\sum_{\ell=0}^{\infty}\frac{\rho^\ell}{d_{\min,i}(\ell)} \;\le\; \frac{1}{d_{\min}},\ \text{for all } i\in V, \tag{7}$$

where the first inequality holds because $\tilde{\mathbf{A}}$ is a row-stochastic matrix. Consequently, we arrive at:

$$|q_i^* - \mu_i^*| \;\le\; C_\rho \cdot \sqrt{(1-\rho)\sum_{\ell=0}^{\infty}\frac{\rho^\ell}{d_{\min,i}(\ell)}},$$

which suggests that the error is small for nodes that have both large in-degrees and are also more distant from nodes with low in-degrees. Hence, we can establish the subsequent corollary.

COROLLARY 1 **(ℓ_∞-Norm Error Bound)**. *Under Assumptions 1 and 2, for any diffusion instance* $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$*, the* ℓ_∞*-norm of the difference between* $\mathbf{q}^*$ *and* $\boldsymbol{\mu}^*$ *can be upper bounded by*

$$\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_\infty \;\le\; C_\rho \cdot \sqrt{\frac{1}{d_{\min}}}. \tag{8}$$

Corollary 1 removes the dependence on the specific network structure from the bound to highlight a worst-case convergence rate as the network expands. Specifically, for a sequence of diffusion instances characterized by an increasing minimum in-degree $d_{\min}$, the maximal deviation shrinks at a rate of $\mathcal{O}\left(\sqrt{1/d_{\min}}\right)$. As $d_{\min}$ approaches infinity, $\boldsymbol{\mu}^*$ is asymptotically equal to $\mathbf{q}^*$. This simplified bound clearly indicates that the FPA scheme can perform better in larger and denser networks. We remark that while the ℓ_∞-norm error bound is intuitively appealing, it can be much looser than the entrywise bound presented in Theorem 1. It relies solely on the minimal in-degree $d_{\min}$, making it overly conservative and sensitive to the isolated outliers. In most real-world networks, the minimal in-degree $d_{\min}$ is often quite small even if its size n is large, limiting the applicability of this bound.

Corollary 2 below addresses this limitation by introducing a bound based on the scaled ℓ_1-norm. Define $r(G) := \max_{i\in V}(\sum_{j=1}^n A_{ij}/\sum_{j=1}^n A_{ji})$ as the largest ratio between out-degree and in-degree.

COROLLARY 2 **(Scaled ℓ_1-Norm Error Bound).** *Under Assumptions 1 and 2, for any diffusion instance $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$, the scaled ℓ_1-norm of the difference can be upper bounded by*

$$\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\le\; C_\rho \cdot \sqrt{\frac{\|\mathcal{C}(G;\rho)\|_1}{n}}. \tag{9}$$

If $r(G) < 1/\rho$, the bound can be further simplified as

$$\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\le\; C_\rho \cdot \sqrt{\frac{1-\rho}{1-\rho r(G)} \cdot \mathcal{D}(G)}. \tag{10}$$

In light of (7), the network-structure-free bound on $\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1$ is also of order $\mathcal{O}\left(\sqrt{1/d_{\min}}\right)$. However, Corollary 2 provides more meaningful bounds. Specifically, (9) bounds the scaled ℓ_1-norm of the error by that of the inverse in-degree centrality $\mathcal{C}(G;\rho)$. More transparent results can be obtained when considering the largest out-to-in-degree ratio $r(G)$. A smaller value of $r(G)$ indicates a more evenly distributed degree structure, with $r(G) \ge 1$ always hold. In poorly conditioned networks, where $r(G)$ is large, substantial weights may apply to nodes with a small in-degree, making the bound approach the worst-case rate $\mathcal{O}\left(\sqrt{1/d_{\min}}\right)$. In contrast, for well-conditioned networks where $r(G) < 1/\rho$, we obtain a bound in (10), that is characterized by the inverse in-degree density $\mathcal{D}(G)$. Unlike $d_{\min}$, which focuses on extreme nodes, $\mathcal{D}(G)$ offers a holistic view by capturing the overall connectivity through the average characteristics of the network. Additionally, $\mathcal{D}(G)$ is also computationally more efficient than both $\mathcal{B}(G;\rho)$ and the inverse in-degree centrality $\mathcal{C}(G;\rho)$, the latter of which requires inverting an n-by-n matrix. As a consequence, $\mathcal{D}(G)$ serves as a more practical and efficient indicator for FPA performance across different networks.

Looking more closely at the impact of $r(G)$ on the bound, we find that the derived upper bound becomes tighter as $r(G)$ decreases, indicating a better performance of the FPA scheme in more

balanced networks. We highlight that the assumption $r(G) < 1/\rho$ for the second part of Corollary 2 is not restrictive in general. Notably, all undirected graphs and balanced directed graphs satisfy that $r(G) = 1 < 1/\rho$. Studies such as Mislove (2009) also validate the balanced nature of social networks in practice. In particular, active agents (i.e., those who create many links) also tend to be popular (i.e., they are the target of many links). This high correlation is generally attributed to the prevalence of reciprocal links in social networks.

We remark on two facts. First, all the aforementioned bounds apply to networks where $d_{\min} > 0$. For any standalone node i with no in-neighbors, the network effect term in (1) is zero, meaning $\mu_i^* = q_i^*$ trivially holds. Including such nodes only tightens the derived bounds. Second, the constant C_ρ decreases with ρ and converges to zero as ρ approaches zero. Therefore, our bounds suggest that FPA works better when ρ, representing the compound effect of network externality and random noise, is small. We emphasize that our theoretical analysis relies on Assumption 2. If this assumption is violated, the bound may lose validity and diverge to infinity. However, in such cases, a constant, such as one or a problem-specific value, can serve as a trivial bound. We defer the investigation of scenarios when this assumption does not hold to our numerical study in Section 5.1.

Operationalizing the FPA solution. On the operational side, the significance of the FPA solution $\boldsymbol{\mu}^*$ lies in the fact that it allows us to reformulate and simplify problem (4). Instead of solving (4) directly, we can replace $\mathbf{q}^*$ with $\boldsymbol{\mu}^*$ and approximate the entire problem (4) as follows:

$$\underset{\mathbf{x}\in\mathcal{X},\boldsymbol{\mu}}{\text{maximize}} \qquad g(\boldsymbol{\mu},\mathbf{x}) \quad \text{s.t. } \boldsymbol{\mu} = \mathbf{h}\Big(\boldsymbol{\mu}\,; G, \mathbf{v}(\mathbf{x}), F_\epsilon(\cdot), \beta\Big), \tag{11}$$

where $\mathbf{h}(\cdot\,; G, \mathbf{v}, F_\epsilon(\cdot), \beta)$ is the adoption evolution operator (AEO) induced by the diffusion instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, which we will formally define using (12) in Section 3.3. The approximate problem (11) offers an explicit formulation by incorporating the FPA scheme as a constraint. This stands in contrast to the implicit variable $\mathbf{q}^*$ in (4), which emerges from a complex stochastic process.

We advocate for the approximate problem (11) for the following reasons: (i) *Theoretical Guarantee.* The FPA scheme is particularly effective for large, dense networks, providing strong theoretical performance guarantees. From a practical standpoint, many real-world networks are large and expanding, making the FPA scheme a promising tool (see Section 5.3). (ii) *Computational Efficiency.* The FPA scheme significantly outperforms agent-based simulation (ABS) in computational efficiency. Fixed-point iteration converges to the FPA solution in linear time (Rheinboldt 1998), whereas ABS requires substantially more computational resources and becomes more cumbersome as network sizes grow. (iii) *Actionable Insights on Network Structure.* The network measure $\mathcal{B}(G;\rho)$ provides an order-optimal characterization of approximation error across arbitrary network structures. Moreover, our proposed metrics, inverse in-degree centrality and inverse in-degree density, are

easy to calculate and serve as accurate FPA performance indicators. (iv) *Closed-Form Expression.* Problem (11) is more tractable than Problem (4), enabling more efficient algorithms tailored for specific problems (see Section 6). (v) *Balanced Precision and Efficiency.* The FPA scheme can integrate with simulation techniques, balancing between precision and computational efficiency. In Section 5.4, we propose a mixture scheme that combines FPA with simulation and empirically demonstrate its effectiveness. By selectively resampling low-degree agents, this approach enhances accuracy with minimal computational overhead, offering scalable, high-precision solutions for different networks.

3.3. Proof Sketch of Theorem 1

In this section, we sketch the proof of Theorem 1, which is our main methodological contribution. The key idea is to construct a deterministic process $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ for a given instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$. We show that $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ closely aligns with the evolution of adoption probability $\{\mathbf{q}(t)\}_{t=0}^{\infty}$.

Specifically, we define $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ as a deterministic dynamic system throughout the time horizon:

$$\mu_i(t) = \begin{cases} q_i(0) & t=0 \\ 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \mu_j(t-1)}{d_i}\right) & t>0 \end{cases}, \quad \text{for all } i \in V. \tag{12}$$

Without loss of generality, we assume $Y_i(0) = 0$ so $q_i(0) = 0$ for all $i \in V$. This is because by Proposition 1, $\mathbf{q}^*$ is unique regardless of the initial distribution $\mathbf{Y}(0)$, and therefore any error bound derived under $\mathbf{Y}(0) = \mathbf{0}$ applies to arbitrary initial distribution of $\mathbf{Y}(0)$. We define $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^n$ as the mapping function that expresses $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ in the form $\boldsymbol{\mu}(t) = \mathbf{h}\big(\boldsymbol{\mu}(t-1)\big)$ for $t \geq 1$. We refer to $\mathbf{h}(\cdot)$ as an adoption evolution operator (AEO) and introduce a family of auxiliary AEOs $\mathcal{H} := \big\{\mathbf{h}_{\boldsymbol{\zeta}}(\cdot) = \mathbf{h}(\cdot) + \boldsymbol{\zeta} : \boldsymbol{\zeta} \in \mathbb{R}^n\big\}$. We proceed by discussing the properties of any AEO $\mathbf{h} \in \mathcal{H}$ and its role in shaping the dynamic system $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$.

PROPOSITION 2 **(Partial Order Preserving, Existence, and Uniqueness).** *Any AEO* $\mathbf{h} \in \mathcal{H}$ *satisfies the following properties* (*i*) *and* (*ii*)*, and the induced dynamic system* $\{\boldsymbol{\nu}(t)\}_{t=0}^{\infty}$ *defined by fixed-point iteration* $\boldsymbol{\nu}(t) = \mathbf{h}\big(\boldsymbol{\nu}(t-1)\big)$ *satisfies the following property* (*iii*)*:*

(*i*) $\mathbf{h}(\mathbf{a}) \leq \mathbf{h}(\mathbf{b})$ *if* $\mathbf{a} \leq \mathbf{b}$.

(*ii*) *There exists a unique fixed-point solution* $\boldsymbol{\nu}^* \in \mathbb{R}^n$ *with* $\mathbf{h}(\boldsymbol{\nu}^*) = \boldsymbol{\nu}^*$.

(*iii*) *For any initial state* $\boldsymbol{\nu}(0)$*, the dynamic system* $\{\boldsymbol{\nu}(t)\}_{t=0}^{\infty}$ *satisfies* $\lim_{t\to\infty} \boldsymbol{\nu}(t) = \boldsymbol{\nu}^*$.

The proof of Proposition 2(i) follows from the definition of $\mathbf{h}(\cdot)$. Proposition 2(ii) and (iii) are consequences of the fact that $\mathbf{h}(\cdot)$ is a contraction mapping. Note that $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ is a special case of the induced dynamic system $\{\boldsymbol{\nu}(t)\}_{t=0}^{\infty}$ when $\boldsymbol{\nu}(0) = \mathbf{q}(0)$. Given these properties, for any diffusion instances under Assumptions 1 and 2, we can always find a well-defined FPA solution $\boldsymbol{\mu}^*$ for limiting adoption probability $\mathbf{q}^*$ by solving the system of equations $\mathbf{h}(\boldsymbol{\mu}) = \boldsymbol{\mu}$.

In this proof, we in fact show a claim stronger than Theorem 1: $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ and $\{\mathbf{q}(t)\}_{t=0}^{\infty}$ are always uniformly close to each other. Taking the limit as $t \to \infty$, we then show Theorem 1. For this,

we face two challenges. The first is the temporal and spatial dependencies in adoptions. An agent's adoption utility is directly shaped by their in-neighbors, and these localized correlations accumulate and propagate across the network over time. The second challenge arises from the nonlinearity of the CDF F_ϵ of a general distribution, which complicates the analytical tracking of adoption state evolution, particularly in characterizing adoption correlations. To address these, our subsequent efforts focus on bounding the spatio-temporal variances and the nonlinear dynamics sequentially.

First, we focus on the local network effect term presented in (1), which represents an average over a set of mutually dependent random variables. To quantify this, we characterize the covariance matrix of adoption states $\mathbf{Y}(t)$ over time in Lemma 1, and then relates it to the variance of local network effect term in Lemma 2.

LEMMA 1 **(Covariance Matrix)**. *Under Assumptions 1 and 2, for any diffusion instance* $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$ *and* $t \geq 1$*, the covariance matrix of adoption states* $\mathbf{Y}(t)$ *can be upper bounded by*

$$\mathbf{\Sigma}(t) \;\leq\; \frac{1}{4}\left[\mathbf{I} + \sum_{\tau=1}^{t-1} \rho^{2\tau} \tilde{\mathbf{A}}^{\tau} \left(\tilde{\mathbf{A}}^{\top}\right)^{\tau}\right].$$

The proof of Lemma 1 establishes the upper bound on $\mathbf{\Sigma}(t)$ through induction. A key step is demonstrating that adoption states $\mathbf{Y}(t)$ remain positively associated (See Definition 2 in Appendix B), which is a stronger form of positive correlation. This property enables the derivation of an iterative upper bound on the covariance matrix, despite the nonlinear dynamics.

To quantify the variance of the network effect, we introduce $\kappa_i(t) := \mathrm{Var}\left(\frac{1}{d_i}\sum_{j\in\mathcal{N}_i} Y_j(t)\right)$. Lemma 2 then provides an upper bound on this variance for each period, based on Lemma 1.

LEMMA 2 **(Variance of Network Effect)**. *Under Assumptions 1 and 2, for any diffusion instance* $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$ *and* $t \geq 1$*, the variance of network effect can be upper bounded by*

$$\boldsymbol{\kappa}(t) \;\leq\; \frac{1}{4}\left(\sum_{\tau=1}^{t} \rho^{2\tau-2} \tilde{\mathbf{A}}^{\tau} \odot \tilde{\mathbf{A}}^{\tau}\right)\mathbf{e} \;\leq\; \frac{1}{4}\left(\mathbf{I} + \sum_{\tau=1}^{t-1} \rho^{2\tau} \tilde{\mathbf{A}}^{\tau}\right)\mathbf{b}.$$

In Lemma 2, the first upper bound provides a detailed measure that characterizes the evolution of the variance $\boldsymbol{\kappa}(t)$. However, this measure is difficult to computed in general. To improve both interpretability and computational efficiency, we introduce the second, looser bound, which is particularly useful in the asymptotic case. Specifically, for $t \geq 1$, we can bound $\boldsymbol{\kappa}(t)$ as follows:

$$\boldsymbol{\kappa}(t) \;\leq\; \frac{1}{4}\left(\sum_{\tau=1}^{\infty} \rho^{2\tau-2} \tilde{\mathbf{A}}^{\tau} \odot \tilde{\mathbf{A}}^{\tau}\right)\mathbf{e} \;=\; \frac{1}{4}\mathcal{B}(G;\rho) \;\leq\; \frac{1}{4}\left[\mathbf{I} + \sum_{\tau=1}^{\infty} \rho^{2\tau} \tilde{\mathbf{A}}^{\tau}\right]\mathbf{b} \;=\; \frac{1}{4}\left(\mathbf{I} - \rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}. \quad (13)$$

Here, we highlight both $(1/4)\cdot\mathcal{B}(G;\rho)$ and $(1/4)\cdot\left(\mathbf{I}-\rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}$ as upper bounds. In the special case where the underlying network structure is a directed ring, both expressions collapse into the same bound. In general, however, each bound offers unique insights and has its own advantages.

On one hand, $(1/4)\cdot\mathcal{B}(G;\rho)$ offers an accurate characterization of how the variation in the network effect propagates. It further enables us to derive a refined upper bound and a matching lower bound, as shown in Section 4, making it an order-optimal measure for arbitrary network structures. However, $\mathcal{B}(G;\rho)$ lacks a simplified expression. Although the sum of infinitely many terms looks like a Neumann series, it is not exactly one due to the Hadamard product in the equation.

On the other hand, $(1/4)\cdot\left(\mathbf{I}-\rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}$ offers a looser bound but provides a closed-form expression and connects the complex measure $\mathcal{B}(G;\rho)$ to centrality metrics resembling our inverse in-degree centrality. This connection helps us to interpret our results more easily. Since $\tilde{\mathbf{A}}$ is row-stochastic, $(1/4)\cdot\mathcal{B}(G;\rho)$ bounds $\boldsymbol{\kappa}(t)$ by (approximately) the weighted sum of inverse in-degrees $\mathbf{b}$. This implies that as the number of in-neighbors increases, the variance decreases. In other words, having more in-neighbors reduces the impact of a neighbor, thus reducing the mutual dependence among adoptions. As time progresses, this upper bound increases, reflecting a discounted contribution from neighbors connected through paths of length $t-1$. As such, while this loose upper bound sacrifices some details, it still offers valuable insights into how stochasticity spreads both spatially and temporally.

With the bound on the variance of network effect, we then move on to bound the nonlinear dynamics. Although the adoption probability $\{\mathbf{q}(t)\}_{t=0}^{\infty}$ lacks a closed-form expression, we expect its transition between consecutive time steps akin to the AEO $\mathbf{h}(\cdot)$.

LEMMA 3 **(Fixed-Point Deviation of Adoption Probability)**. *Under Assumptions 1 and 2, for any diffusion instance* $\left(G,\mathbf{v},F_\epsilon(\cdot),\beta\right)$ *and* $t\geq 1$*, we have*

$$\left|\mathbf{h}\big(\mathbf{q}(t-1)\big)-\mathbf{q}(t)\right|_{\text{ew}} \;\leq\; \frac{\rho}{2}\left[\mathcal{B}(G;\rho)\right]_{\text{ew}}^{\frac{1}{2}} \leq \frac{\rho}{2}\left[\left(\mathbf{I}-\rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}\right]_{\text{ew}}^{\frac{1}{2}}.$$

Building on Lemma 2, Lemma 3 essentially connects the transitions of $\{\mathbf{q}(t)\}_{t=0}^{\infty}$ and $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$, offering a one-step guarantee for them. We then use a "fixed-point sandwich" technique to prove the results in Theorem 1. Specifically, let $\boldsymbol{\delta}:=\frac{\rho}{2}[\mathcal{B}(G;\rho)]_{\text{ew}}^{\frac{1}{2}}$ denote the absolute deviation. We define a lower bound system $\{\underline{\boldsymbol{\mu}}(t)\}_{t=0}^{\infty}$ and an upper bound system $\{\overline{\boldsymbol{\mu}}(t)\}_{t=0}^{\infty}$ based on this deviation as follows:

$$\underline{\mu}_i(t)=\begin{cases} q_i(0) & t=0\\ 1-F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i}\underline{\mu}_j(t-1)}{d_i}\right)-\delta_i & t>0\end{cases},\quad \text{for all } i\in V,$$

$$\overline{\mu}_i(t)=\begin{cases} q_i(0) & t=0\\ 1-F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i}\overline{\mu}_j(t-1)}{d_i}\right)+\delta_i & t>0\end{cases},\quad \text{for all } i\in V.$$

Using auxiliary AEOs, these two systems can be expressed as fixed-point iterations: $\underline{\boldsymbol{\mu}}(t)=\mathbf{h}_{-\boldsymbol{\delta}}(\underline{\boldsymbol{\mu}}(t-1))$ and $\overline{\boldsymbol{\mu}}(t)=\mathbf{h}_{\boldsymbol{\delta}}(\overline{\boldsymbol{\mu}}(t-1))$, with $\underline{\boldsymbol{\mu}}^*$ and $\overline{\boldsymbol{\mu}}^*$ being the respective fixed-point solutions. Note that $\mathbf{h}_{-\boldsymbol{\delta}},\mathbf{h}_{\boldsymbol{\delta}}\in\mathcal{H}$. Then, we leverage these two fixed-point iterations to sandwich both $\{\mathbf{q}(t)\}_{t=0}^{\infty}$ and

$\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ and show that both $\underline{\boldsymbol{\mu}}(t) \le \mathbf{q}(t) \le \overline{\boldsymbol{\mu}}(t)$ and $\underline{\boldsymbol{\mu}}(t) \le \boldsymbol{\mu}(t) \le \overline{\boldsymbol{\mu}}(t)$ hold. Further, one can argue that $\mathbf{0} \le \overline{\boldsymbol{\mu}}(t) - \underline{\boldsymbol{\mu}}(t) \le 2\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1}\boldsymbol{\delta}$. Therefore, $|\mathbf{q}(t) - \boldsymbol{\mu}(t)|_{\text{ew}} \le 2\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1}\boldsymbol{\delta}$. Sending t to infinity, the result of Theorem 1 holds. For details, please refer to Appendix B.3.

We highlight that, while for clarity in Theorem 1 we frame our theoretical result in terms of the limiting behavior (i.e., $\boldsymbol{\mu}^*$ and $\mathbf{q}^*$), our arguments here show that these bounds apply to the entire diffusion trajectory. In particular, the fixed-point iteration $\boldsymbol{\mu}(t)$ also approximates the transient adoption probability $\mathbf{q}(t)$ even when t is small, and the approximation error is naturally bounded by the gap between their limiting behaviors, thereby extending the result to the full diffusion process.

4. Order-Optimal Error Bounds for the FPA Scheme

In this section, we delve deeper into the error bound of the FPA scheme. By introducing an additional mild assumption on the noise distribution, we derive a tighter upper bound and present a matching lower bound of the same order, thus closing the gap in our analysis. For the subsequent analysis, we will proceed under Assumption 3.

ASSUMPTION 3 **(Stronger Smoothness Condition)**. *The random noise $\epsilon_i(t)$ has a differentiable probability density function (PDF) $f_\epsilon(\cdot)$ with its derivative upper bounded by $|f'_\epsilon(\cdot)| \le L_f$.*

This assumption mainly requires the smoothness of the PDF $f_\epsilon(\cdot)$. It is worth noting that this assumption is fairly mild, given that many commonly used distributions inherently exhibit high degrees of differentiability, including but not limited to the normal and logistic distributions.

4.1. Improved Upper Bounds

Recall that Theorem 1 establishes an upper bound for the approximation error at the order of $\left[\mathcal{B}(G;\rho)\right]_{\text{ew}}^{\frac{1}{2}}$. Under Assumption 3, Theorem 2 refines this upper bound, improving it to a lower order of $\mathcal{B}(G;\rho)$. Define the constant $\tilde{C} := \frac{L_f\beta^2}{4(1-\rho)^2}$, which increases with both ρ and L_f.

THEOREM 2 **(Improved Entrywise Error Bound of the FPA Scheme)**. *Under Assumptions 1, 2 and 3, for any diffusion instance $\left(G, \mathbf{v}, F_\epsilon(\cdot), \beta\right)$, we have*

$$|\mathbf{q}^* - \boldsymbol{\mu}^*|_{\text{ew}} \le \frac{L_f\beta^2}{4}\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1}\mathcal{B}(G;\rho) \;\le\; \tilde{C}\cdot\mathcal{C}(G;\rho) \quad \textit{and} \quad \|\mathbf{q}^* - \boldsymbol{\mu}^*\|_\infty \;\le\; \tilde{C}\cdot\frac{1}{d_{\min}}. \tag{14}$$

We observe that both $\mathcal{B}(G;\rho)$ and the inverse in-degree centrality $\mathcal{C}(G;\rho)$ remain crucial in the improved bound. Unlike Theorem 1, the dependency of the bound on these measures is now linear. In light of the refined bound (14), we also sharpen the scaled ℓ_1-norm of the approximation error.

COROLLARY 3 **(Improved Scaled ℓ_1-Norm Bound)**. *Under Assumptions 1, 2 and 3, for any diffusion instance $\left(G, \mathbf{v}, F_\epsilon(\cdot), \beta\right)$ with $r(G) < 1/\rho$, we have*

$$\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\le\; \frac{(1-\rho)\tilde{C}}{1-\rho r(G)}\cdot\mathcal{D}(G). \tag{15}$$

The proof of Theorem 2 largely parallels that of Theorem 1, with the key difference being the use of a second-order Taylor expansion of F_ϵ, enabled by Assumption 3, to bound $|\mathbf{h}(\mathbf{q}(t-1)) - \mathbf{q}(t)|_{\text{ew}}$.

LEMMA 4 **(Improved Fixed-Point Deviation of Adoption Probability)**. *Under Assumptions 1, 2 and 3, for any diffusion instance* $\left(G, \mathbf{v}, F_\epsilon(\cdot), \beta\right)$ *and* $t \geq 1$*, we have*

$$\left|\mathbf{h}(\mathbf{q}(t-1)) - \mathbf{q}(t)\right|_{\text{ew}} \;\leq\; \frac{L_f \beta^2}{8} \mathcal{B}(G;\rho) \;\leq\; \frac{L_f \beta^2}{8} \left(\mathbf{I} - \rho^2 \tilde{\mathbf{A}}\right)^{-1} \mathbf{b}.$$

4.2. A Matching Lower Bound

Our results in Section 4.1 highlight a linear dependence of FPA's error upper bounds on network measures such as $\mathcal{B}(G;\rho)$, $\mathcal{C}(G;\rho)$, $1/d_{\min}$, and $\mathcal{D}(G)$. We now establish matching lower bounds of the same order, suggesting that these bounds are order-optimal.

THEOREM 3 **(Entrywise Lower Bound of the Error)**. *For any given network* G*, there exists a diffusion instance* $\left(G, \mathbf{v}, F_\epsilon(\cdot), \beta\right)$ *satisfying Assumptions 1, 2 and 3 such that*

$$|\mathbf{q}^* - \boldsymbol{\mu}^*|_{ew} \;\geq\; \hat{C} \left(\mathbf{I} - \rho_\ell \tilde{\mathbf{A}}\right)^{-1} \mathcal{B}(G;\rho_\ell), \tag{16}$$

where $\hat{C}$ *depends on* $\mathbf{v}$*,* $F_\epsilon(\cdot)$ *and* β*.*

Notably, the constants $\tilde{C}$ and ρ in (14) depend on the distribution of random error, particularly the upper continuity bounds. The constants in the matching lower bound, i.e., $\hat{C}$ and ρ_ℓ, instead depends on the lower continuity bounds. Specifically, to establish the lower bound (16), for any $\xi > 0$, we construct a diffusion instance where (i) each agent $i \in V$ has the same intrinsic value $v = -\beta - \xi$, and (ii) the random noise follows $\epsilon_i(t) \sim \text{Logistic}(0,1)$. A key consequence is that for any node $i \in V$,

$$v_i + \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} \;\in\; [v, v+\beta] = [-\beta - \xi, -\xi],$$

so we can set ℓ and ℓ_f as arbitrary values satisfying $|F_\epsilon(x) - F_\epsilon(y)| \geq \ell |x - y|$ and $|f'_\epsilon(x)| \geq \ell_f$, respectively, for all x, y in $[-\beta - \xi, -\xi]$. These bounds allow us to reverse the chain of inequalities in the analysis of the upper bounds, eventually leading to the establishment of (16), with $\hat{C} = \ell_f \beta^2 [1 - F_\epsilon(\xi + \beta)] F_\epsilon(\xi)/2$. Furthermore, when the underlying graph is a regular graph, we can also establish the matching lower bounds with regard to other proposed network metrics, i.e., $\mathcal{C}(G;\rho)$, $1/d_{\min}$, and $\mathcal{D}(G)$.We comment that the constants in the lower and upper bounds do not match and we relegate the analysis regarding the (sub-)optimality of the constants to future research.

5. Numerical Experiments

In this section, we present numerical studies to validate our FPA scheme in different scenarios, including the 10-node example (i.e., Figure 1a), large-scale random networks, and real-world networks. Our results reveal several key insights. First, the FPA scheme consistently achieves superior

performance, even for small and sparse networks. Second, the inverse in-degree centrality $\mathcal{C}(G;\rho)$ and the inverse in-degree intensity $\mathcal{D}(G)$ emerge as strong indivators of FPA performance. Third, in terms of computational efficiency, the FPA scheme significantly outperforms alternative simulation approaches. Lastly, a heuristic mixture scheme, which combines the FPA scheme with resampling, further improves approximation performance with only a slight increase in computational overhead.

A fundamental challenge in measuring FPA performance is that the limiting adoption probability $\mathbf{q}^*$ is generally unknown. As outlined in Section 2, deriving $\mathbf{q}^*$ requires solving the stationary distribution of a large-scale MC, which is typically computationally infeasible[4]. As a practical workaround, we resort to agent-based simulation (ABS) approach over a long time horizon to estimate the ground-truth $\mathbf{q}^*$. See Appendix E.3 for details. Unless otherwise specified, we adopt the following experimental settings: The limiting adoption probability $\mathbf{q}^*$ is estimated by ABS, while the FPA solution $\boldsymbol{\mu}^*$ is obtained through fixed-point iteration, with an initial value $\boldsymbol{\mu}(0)=\mathbf{0}$ and a convergence criterion of 10^{-5}. The noise distribution follows $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0,1)$ and the network effect intensity is set to $\beta=3.5$, so $\rho=0.875$. With this relatively high value of ρ, our experiments aim to explore near-worst-case scenarios, offering a robust evaluation of the FPA scheme. To quantify performance for a specific diffusion instance, we use the mean absolute percentage error (MAPE) across all agents, defined as $\text{MAPE}=\frac{1}{n}\sum_{i\in V}\frac{|\mu_i^*-q_i^*|}{q_i^*}\times 100\%$. This self-normalized measure, instead of the absolute errors used in the theoretical analysis, allows us to focus on relative errors, making the results more interpretable and facilitating comparisons across different adoption probabilities.

5.1. Revisiting the Motivating Example

In this subsection, we revisit the 10-node motivating example introduced in Section 3.1. We focus on two aspects: the role of our centrality measure as a node-level metric for evaluating FPA performance, and a thorough examination of Assumption 2, which serves as a sufficient condition for all our theoretical results. For experiments in this subsection, $\mathbf{q}^*$ is achieved by solving the exact stationary distribution of the MC due to the manageable size.

5.1.1. The role of inverse in-degree centrality. As highlighted in our theorems, the upper bound of the approximation error is linked to the inverse in-degree centrality $\mathcal{C}(G;\rho)$. In Figure 2, we juxtapose the absolute approximation error $|q_i^*-\mu_i^*|$ with the metric $\mathcal{C}_i(G,\rho)$ for all 10 nodes. The results reveal a clear positive correlation, emphasizing the role of inverse in-degree centrality as a sharp node-level indicator for evaluating FPA performance.

[4] For some highly structured symmetric networks (such as star networks and complete networks), solving the stationary distribution is tractable. See Appendix E.2 for details.

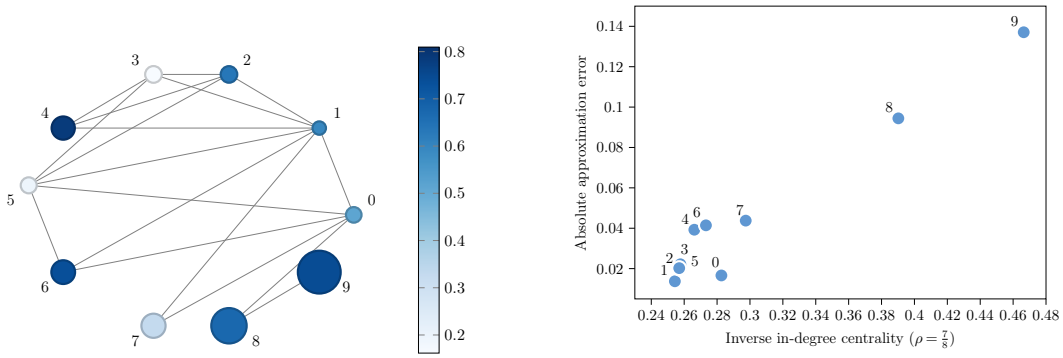
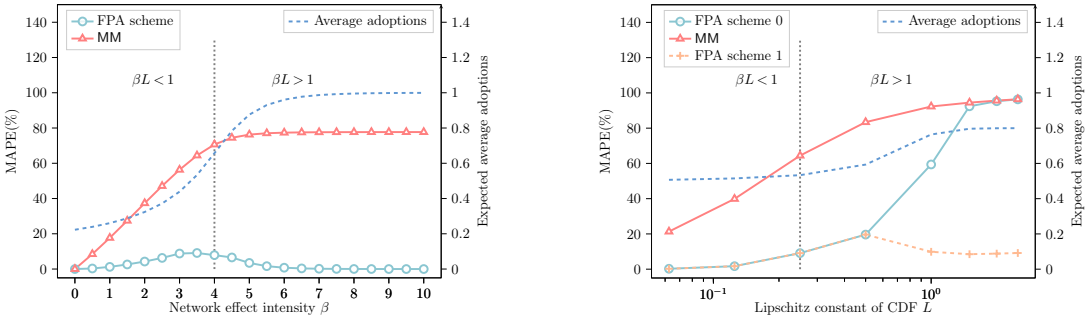


Figure 2 **Analysis of the 10-node example instance. Left: Reproduction of Figure 1a for reference; Right: Illustration of the relationship between the absolute approximation error and inverse in-degree centrality.**

5.1.2. Discussions on Assumption 2. Although assumptions similar to Assumption 2 are common in the network literature, its implications for the FPA scheme deserve further exploration. The parameter ρ has two elements, namely the network effect intensity β and the Lipschitz constant L of $F_\epsilon(\cdot)$. We conduct experiments by varying each of these two components, benchmarking against the misspecified model where the network effect is not incorporated (See MM in Section 3.1). Figure 3 shows the MAPE of the FPA scheme with different values of β and L.



(a) Sensitivity regarding network effect intensity

(b) Sensitivity regarding noise distribution

Figure 3 **Sensitivity of the approximation error against parameter ρ. In the right subfigure, the horizontal axis is in log scale. The FPA scheme 0 (1, resp.) represents the FPA solution initialized with $\mu(0) = 0$ ($\mu(0) = 1$, resp.).**

(i) *Sensitivity analysis of network effect intensity*. We vary β from 0 to 10 and keep other parameters fixed. Given $\epsilon_i \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0,1)$ with $L = 1/4$, our experiments cover both cases when Assumption 2 is satisfied and violated. As shown in Figure 3a, the MAPE first increases with β, but at a notably slower rate compared to MM. When $\rho = \beta L > 1$, the MAPE gradually declines to 0.

For $\rho < 1$, it is not surprising that the FPA scheme performs exceptionally well when β is close to 0 (i.e., the network effect is weak). As β increases, the network effect becomes more influential, slightly degrading FPA performance. This dependence is reflected in the constants C_ρ and $\tilde{C}$ in our

theoretical results. Nevertheless, even at $\beta = 3.5$, where the MAPE peaks at 9.11%, FPA performance remains commendable, substantially lower than that of MM (64.42%). This underscores the robustness of the FPA scheme, even in the presence of strong network effects in a small network.

When $\rho > 1$, we observe an intriguing trend: the MAPE decreases as β increases. This is because L is not uniformly tight for $F_\epsilon(\cdot)$. With large β, the network effects heavily influence user behavior, pushing adoption probabilities close to 1 for many agents. Thus, the nominal utility in the FPA scheme, given by $v_i + \beta \frac{\sum_{j\in\mathcal{N}_i} \mu_j(t-1)}{d_i}$, gravitates towards the flat areas of the CDF, where the effective Lipschitz constant is much smaller than L. Even if the nominal utility occasionally falls into regions with larger Lipschitz constants, thereby violating Assumption 2, the FPA scheme remains resilient.

(ii) *Sensitivity analysis of noise distribution*. The experiment above may lead readers to conceive that Assumption 2 is conservative. However, we demonstrate that this is not the case by varying L. Particularly, we assess FPA performance for Logistic distribution with different parameters, where $\epsilon_i \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0, s)$ and s ranges from 0.0625 to 2.5. The associated Lipschitz constant is given by $L = \frac{1}{4s}$, and we hold other parameters fixed. In Figure 3b, we observe a continuous increase in MAPE as L increases. Notably, when $\rho > 1$, FPA performance deteriorates drastically, approaching that of MM. This deterioration results from the violation of Proposition 2, e.g., there may be multiple solutions to the FPA scheme. In particular, the FPA solution initialized at $\boldsymbol{\mu}(0) = \mathbf{0}$ diverges from $\mathbf{q}^*$. To offer a more comprehensive view, we also present an additional FPA solution initialized at $\boldsymbol{\mu}(0) = \mathbf{1}$ in Figure 3b. These two solutions exhibit divergent performance, with the latter significantly outperforming the former when $\rho > 1$. We remark that other FPA solutions could exist with different initial values, and it is challenging to determine which solution will perform best *a priori*.

Upon closer examination, a key difference emerges between these two scenarios. In the first scenario, the intrinsic utility $\mathbf{v}$ retains its relative position within the noise distribution. In contrast, in the second scenario, when s is small, the relative placement of v_i varies significantly across agents, potentially lying far to the left or right of the noise distribution. As a result, in the latter case, some agents are strongly inclined to adopt, while others tend toward non-adoption. When Assumption 2 is violated in this context, the FPA scheme has multiple solutions, complicating its application. In short, the impacts of violating Assumption 2 on FPA performance are multifaceted and depend on specific circumstances. We leave further explorations of these phenomena to future studies.

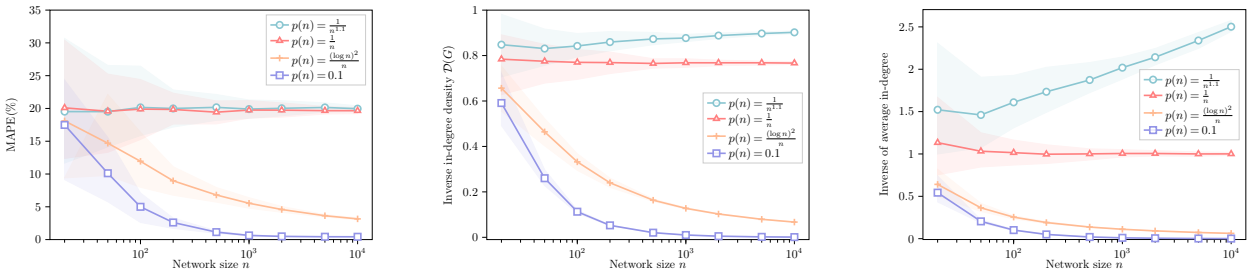
5.2. Random Networks

In this section, we evaluate FPA performance across various well-studied random networks, focusing on Erdős-Rényi networks of different sizes and densities. We also investigate FPA performance on power-law networks, varying the exponent and degree correlation and include these supplementary results in Appendix E.4. To ensure robustness, we conduct 50 repetitions for each combination

of random network parameters. In the following, we consider a sequence of directed Erdős-Rényi networks, each denoted by $G(n, p(n))$, where n represents the network size and $p(n)$ represents the network density—the probability that an edge between two nodes exists. The presence of edges is assumed to be independent. We conduct experiments from two perspectives: the sensitivity of FPA performance to network structures and the computational efficiency of the FPA scheme.

5.2.1. FPA performance with regard to network structure. We assess FPA performance by varying the size and density of Erdős-Rényi networks, respectively.

(i) *Network size*. We vary the network size n from 20 to 10,000, and select densities $p(n)$ from the set $\{\frac{1}{n^{1.1}}, \frac{1}{n}, \frac{(\log n)^2}{n}, 0.1\}$. These values are chosen based on the critical ranges of $p(n)$ identified in random graph literatures. Further details on the properties of Erdős-Rényi networks are provided in Appendix E.5 (see also, Huang et al. 2022). Our theoretical results, as discussed in Sections 3 and 4, assume $d_{\min} > 0$. However, we may encounter standalone nodes with in-degree 0 in Erdős-Rényi networks. These nodes, which do not receive influence but can still exert influence (i.e., their out-degree can be positive), are perfectly approximated in the FPA solution. To align with the theoretical framework, we exclude such standalone nodes from our analysis in this subsection. However, in other sections where we evaluate the overall FPA performance across different network configurations, we include these nodes to ensure a more comprehensive assessment. In practice, social networks are typically integrated systems, so excluding standalone nodes may not reflect real-world scenarios.

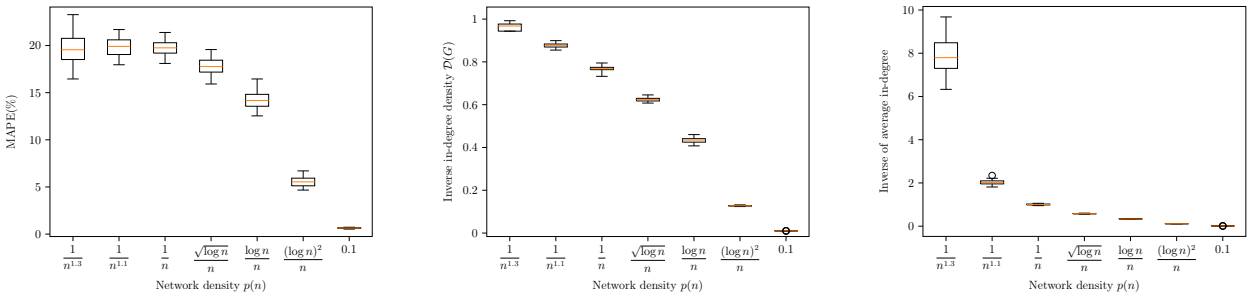


(a) MAPE against n (b) $\mathcal{D}(G)$ against n (c) Inverse of average in-degree against n

Figure 4 FPA performance on Erdős-Rényi networks of different network sizes. All horizontal axes are in the log scale. Shaded areas represent the 95% confidence interval.

In Figure 4a, we examine how MAPE varies with network size n. For all tested cases, the MAPE either decreases or remains stable as n grows, with a more pronounced decrease when $p(n)$ is large. Figures 4b and 4c present two metrics—the inverse in-degree density $\mathcal{D}(G)$ and the inverse of average in-degree. As n increases, we find that $\mathcal{D}(G)$ trends similarly to MAPE, indicating that it is a strong indicator of FPA performance. In contrast, the average in-degree, which traditionally indicates density, shows weaker explanatory power for this trend.

(ii) *Network density*: We fix a medium network size of $n = 1,000$ and focus on various densities $p(n) \in \left\{ \frac{1}{n^{1.3}}, \frac{1}{n^{1.1}}, \frac{1}{n}, \frac{\sqrt{\log n}}{n}, \frac{\log n}{n}, \frac{(\log n)^2}{n}, 0.1 \right\}$. In Figure 5a, the MAPE decreases as the network gets denser, reflecting a trend that aligns with Figure 4a and our theoretical results.



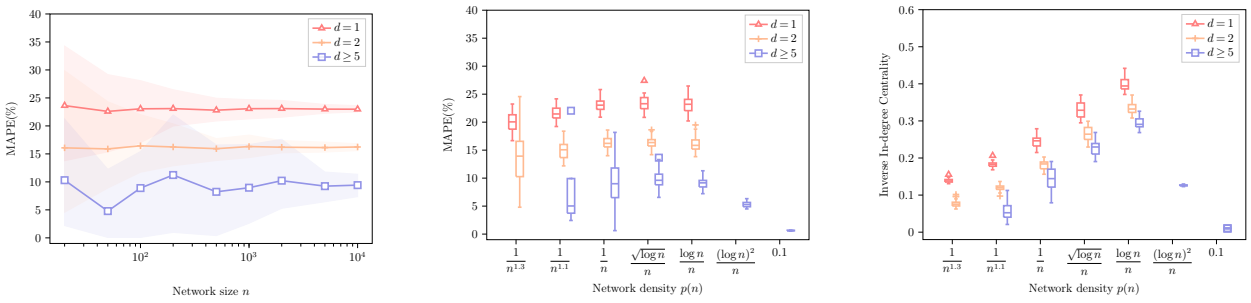
(a) MAPE against $p(n)$ (b) $\mathcal{D}(G)$ against $p(n)$ (c) Inverse of average in-degree against $p(n)$

Figure 5 FPA performance on Erdős-Rényi networks of different network densities.

Recall that we set ρ close to 1 to demonstrate the near-worst-case performance. Even under such a setting, the FPA scheme performs exceptionally well. Figures 5b and 5c further show two network metrics, $\mathcal{D}(G)$ and the inverse of average in-degree, respectively. The trends largely mirror those seen in Figures 4b and 4c, again confirming that $\mathcal{D}(G)$ is an informative indicator of FPA performance.

We observe that the enhanced performance of dense networks can also be partially attributed to the largest out-in-degree ratio $r(G)$. As highlighted in Corollary 2, the upper bound for approximation error increases as $r(G)$ increases. In dense Erdős-Rényi networks, both in-degrees and out-degrees of nodes tend to cluster around the mean value. This contributes to the density of the network while also making it more balanced. To further explore the role of network imbalance, we conduct an extensive analysis with power-law networks, constructing in-degree and out-degree sequences with different correlations (see Appendix E.4).

5.2.2. FPA performance for low in-degree agents. We complement our analysis by singling out the agents with low in-degrees, who, as illustrated by both numerical and theoretical analyses, exhibit poor performance in the FPA scheme. We specifically analyze the MAPE for nodes with $d = 1, 2$, and those with $d \geq 5$ for comparison, visualized in Figure 6, across various network sizes and densities. Key takeaways include the following: Agents with an in-degree of 1 exhibit larger MAPE compared to those with larger in-degrees, as shown in Figures 4a and 5a. These results reinforce our node-level theoretical analyses (Theorems 1 and 2). Additionally, for agents with the same in-degree, we find that their MAPEs remain relatively stable as n increases, though there is a slight upward trend as $p(n)$ increases. This trend can be attributed to the influence of more distant neighbors, highlighting the importance of capturing the overall network structure and connectivity. Furthermore, we find that this structural information is well represented by the inverse in-degree



(a) MAPE against n (b) MAPE against $p(n)$ (c) $\mathcal{C}(G;\rho)$ against $p(n)$

Figure 6 Degree-level FPA performance. In the left subfigure, $p(n)$ is fixed to be $\frac{1}{n}$ and the horizontal axis is in log scale. Shaded areas represent the 95% confidence interval. In the middle subfigure, n is fixed to be $1,000$.

centrality, as shown in Figure 6c. The patterns in the centrality measure closely match those of MAPE, and this alignment would be nearly perfect if we used the mean absolute error instead. However, we have omitted those results for brevity.

5.2.3. Computational efficiency of the FPA scheme. We evaluate the computational efficiency by comparing the CPU time required to compute $\boldsymbol{\mu}^*$ with that needed to estimate $\mathbf{q}^*$ using simulation methods. To provide a comprehensive evaluation of FPA performance, we use two simulation methods as benchmarks to estimate $\mathbf{q}^*$. The first is the naïve ABS method. This method simulates the evolution of adoption states based on the agent behavior model defined in (1) and estimates $\mathbf{q}^*$ by averaging adoption states after the warm-up period, as described in the first equation of (3). To ensure a fair comparison, we report the simulation time for the naïve ABS method once its real-time MAPE falls below the value achieved by the FPA scheme. While the naïve ABS method is a natural approach for simulating limiting adoption probability, it can be time-consuming and does not fully capture the performance of advanced simulation techniques. To address this, we also propose an accelerated ABS (A-ABS) method, which improves simulation efficiency by leveraging the analytical form of the behavior model. A-ABS estimates the adoption probability by averaging conditional probabilities, as described in the second equation of (3). More details on our simulation benchmarks are provided in Appendix E.3.

We fix the network density at $n(p)=0.1$ and vary n from 20 to 10,000 to test the runtime of different methods. As shown in Table 1, the runtime for both simulation methods and the FPA scheme increases with n. However, the FPA scheme consistently outperforms both simulation methods by a substantial margin, with the performance gap widening for larger networks. For instance, when $n=10,000$, approximately 40 minutes are required for the naïve ABS method and 6.5 minutes for the A-ABS method to achieve the same performance as the FPA scheme, which completes the task in just 2.3 seconds. This illustrates that in such a large and dense network, the FPA scheme is about 1,027 times and 170 times more efficient than the naïve ABS and A-ABS methods, respectively.

Table 1 The CPU time required for simulation methods and the FPA scheme.

Network size n	20	50	100	200	500	1,000	2,000	5,000	10,000
Naïve ABS time (s)	0.168	0.258	0.530	2.415	18.979	97.452	286.718	1315.301	2366.879
A-ABS time (s)	0.152	0.285	0.512	0.996	2.380	5.456	15.083	93.235	391.357
FPA time (s)	0.002	0.004	0.007	0.018	0.044	0.102	0.230	1.027	2.304

In conclusion, the FPA scheme offers considerable advantages in computational efficiency across all the tested scenarios without a significant compromise in accuracy, which implies its potential to effectively characterize the diffusion process for a large variety of social networks.

5.3. Real-World Networks

Our numerical experiments in Section 5.2 focus on the random networks, which may not capture real world phenomena (e.g., see Jackson 2010). To evaluate the FPA scheme in more realistic settings, we test it on real-world networks from the Network Repository (Rossi and Ahmed 2015). Specifically, we select five social friendship networks extracted from Facebook, where nodes represent individuals and edges denote friendship ties. A summary of the networks and their results is provided in Table 2. For raw data and additional statistics, readers can refer to the Network Repository website[5].

Table 2 Experiment results for real-world networks.

Instance	n	$d_{\min}$	$d_{\max}$	$\bar{d}$	$\mathcal{D}(G)$	MAPE(%)	Computation time (s)		
							naïve ABS	A-ABS	FPA
Caltech36	770	1	248	43.2623	0.1108	3.48	4.5335	1.5075	0.0636
Reed98	963	1	313	39.0696	0.0962	3.14	5.6228	1.9044	0.0623
Haverford76	1,447	1	375	82.3621	0.0427	1.59	23.2347	2.6996	0.1009
Simmons81	1,519	1	300	43.4338	0.0857	2.85	11.9185	3.4207	0.1426
Amherst41	2,236	1	467	81.3542	0.0488	1.71	35.6749	4.7757	0.1846

We emphasize three key observations from these experiments. First, the FPA scheme performs exceptionally well, achieving a maximum MAPE of just 3.48% across all evaluated networks, indicating both accuracy and reliability for real-world applications. Second, the FPA scheme significantly outperforms the naïve ABS method in terms of computational time, with a speed-up factor ranging from 70 to 230. Even compared to the A-ABS method, the FPA scheme is 23 to 30 times faster, showcasing its computational efficiency and scalability. Third, among various metrics such as network size n, minimal in-degree $d_{\min}$, maximal in-degree $d_{\max}$, and average in-degrees $\bar{y}$, the inverse in-degree density $\mathcal{D}(G)$ stands out as the most reliable indicator of FPA performance measured by MAPE.

Figure 7 presents a comprehensive overview of the relationship between FPA performance and the inverse in-degree density across different families of networks. The figure reveals a clear positive

[5] See https://networkrepository.com/networks.php.

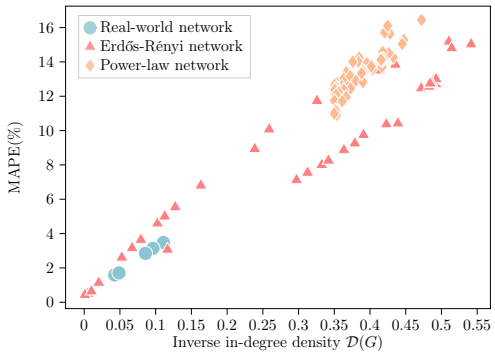


Figure 7 **FPA performance against the inverse in-degree density for different networks. Each point for the random networks represents the average values of the same parameter pair across all repetitions.**

correlation between MAPE and $\mathcal{D}(G)$, reinforcing that $\mathcal{D}(G)$ is not just an upper bound of performance but also a reliable and easy-to-compute metric to measure FPA performance. Interestingly, real-world networks generally exhibit lower $\mathcal{D}(G)$ values and MAPE for the FPA scheme compared to most random networks. This suggests that random networks may not fully capture the characteristics of social networks in this approximation setting, further emphasizing the practical relevance of our FPA scheme for arbitrary network structures.

5.4. A Heuristic Mixture of the FPA and Simulation Scheme

While the FPA scheme improves time efficiency in network diffusion characterization, it sacrifices precision, particularly for low in-degree agents, as shown both theoretically and empirically in Section 5.2.3. This limitation stems from $\boldsymbol{\mu}$ being a deterministic approximation, which loses accuracy under nonlinear transformations. Even if the FPA scheme accurately approximates the adoption probabilities of an agent's in-neighbors, it may not be accurate for the agent itself due to the deterministic nature. To mitigate this, we propose a heuristic mixture scheme that combines resampling with the FPA scheme, improving accuracy with a slight increase in computational overhead.

In this mixture scheme, we first derive the FPA solution $\boldsymbol{\mu}^*$. Next, we resample the adoption states of agent i as independent Bernoulli variables $\tilde{Y}_i$ with probability μ_i^*. The adoption probability for agent i can then be refined through simulation by evaluating the following expected value:

$$\tilde{q}_i = \mathbb{E}\left[1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i}\tilde{Y}_j}{d_i}\right)\right]. \tag{17}$$

In terms of precision, the mixture scheme is particularly effective for low in-degree agents whose neighbors have large in-degrees. In such cases, the FPA scheme accurately approximates the neighbors' adoption states, while the resampling introduces the necessary stochasticity to refine the adoption probability for low in-degree agents. In terms of efficiency, the mixture scheme remains scalable, as the resampling can be parallelized without rerunning the full simulation. Moreover, it

allows for selective resampling of only low-degree agents, avoiding unnecessary computation for all agents. This balance ensures improved accuracy while maintaining minimal computational overhead.

When implementing this mixture scheme, two key factors must be considered: (1) determining which agents should undergo resampling, and (2) selecting an appropriate sample size to ensure accurate estimation of $\tilde{\mathbf{q}}$ as shown in (17). To illustrate its performance, we evaluate the MAPE of the mixture scheme on two real-world networks, *Caltech36* and *Amherst41*, as introduced in Section 5.3. We vary the cutoff in-degrees, applying the mixture scheme only to agents with in-degrees below such value, ranging from 0 to 200, and consider sample sizes of $\{100, 1000\}$. Notably, when the cutoff degree is set to 0, the mixture scheme reduces to the FPA scheme alone.

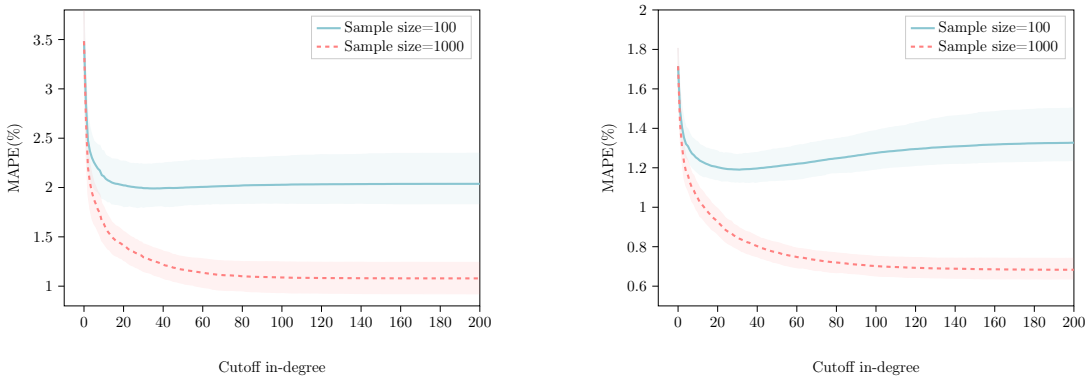


Figure 8 **Performance of the mixture scheme under different conditions. Left: *Caltech36*; Right: *Amherst41*. Shaded areas represent the 95% confidence interval.**

In Figure 8, we present the performance of the mixture scheme under various conditions. The results show that the mixture scheme significantly outperforms the FPA scheme across all scenarios, with the most notable improvements observed for low in-degree agents. This is reflected in the sharp decrease in MAPE when the cutoff in-degree is small. Moreover, larger sample sizes, which provide more accurate estimations of (17), lead to better performance and greater flexibility in selecting the cutoff in-degree. With a sample size of 100, MAPE initially decreases as the cutoff in-degree increases, reaching a minimum around 20, before rising again in both instances. This suggests that coarse estimations of $\tilde{\mathbf{q}}$ work well for low in-degree agents but can negatively affect high in-degree agents, who are already well-approximated by the FPA scheme. This emphasizes the importance of carefully selecting the cutoff in-degree. When the sample size is increased to 1,000, which provides a more accurate evaluation of $\tilde{\mathbf{q}}$, MAPE decreases almost monotonically as the cutoff in-degree increases, eventually reducing to around 1% for both instances. This shows that with sufficient computational resources, the cutoff can be adjusted more flexibly.

Notably, the warm-up period for the ABS methods is set to 1,000. By parallelizing computations and focusing only on low in-degree agents, the computation time for the mixture scheme can be

significantly smaller than the warm-up period of the ABS methods. However, it is important to note that the MAPE of the mixture scheme does not reach zero. This is because, although resampling introduces stochasticity, the samples are independent, whereas the actual adoption states of different agents are positively correlated. Since we cannot capture the full joint distribution of adoption states, this limitation is inherent to this mixture scheme.

In Figure 9, we present a comprehensive comparison of the mixture scheme across various synthetic and real-world networks, alongside the FPA scheme. The mixture scheme consistently shows significant improvements, particularly in cases where the FPA scheme performs poorly. On average, using a cutoff in-degree of 20 and a sample size of 100, the mixture scheme reduces the MAPE by 47.08%, while a cutoff in-degree of 200 and a sample size of 1,000 achieves a reduction of 58.94%. In conclusion, the resampling approach effectively complements the FPA scheme. In practice, the

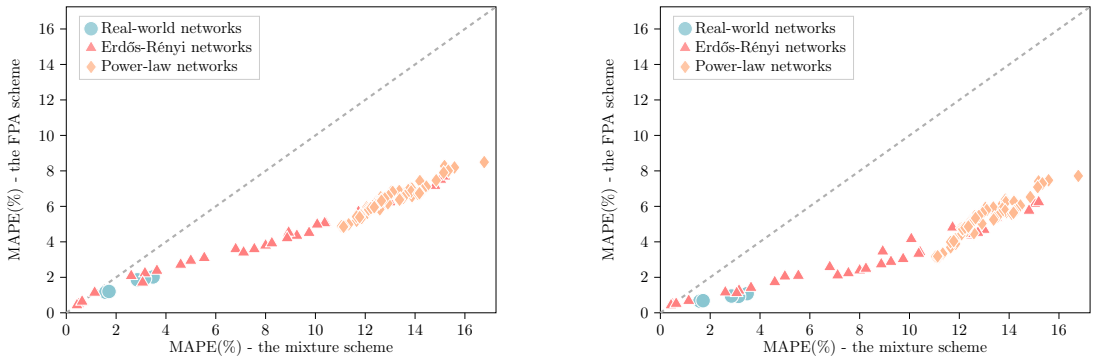


Figure 9 Performance comparison between the mixture scheme and the FPA scheme. Left: *Caltech36*; Right: *Amherst41*. The grey dotted line represents $y = x$, indicating equal performance of two schemes.

cutoff in-degree and sample size can be adjusted to balance precision and efficiency.

6. Applications of the Fixed-Point Approximation Scheme

The FPA scheme can be applied to many classical operational decision problems involving network diffusion. In this section, we consider two notable applications: influence maximization (IM) problem in network analysis, and optimal pricing in revenue management. Hereafter, we confine our analysis of the optimization problems to a given instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$.

We first analyze how the approximation error of the FPA scheme translates into the optimality gap of the optimization problem. Considering the general optimization problem (4) and its approximate formulation (11), we define the regret for a platform decision $\mathbf{x}$ as the difference between the optimal objective value and the objective value under $\mathbf{x}$. Formally, the regret is given by:

$$\text{Regret}(\mathbf{x}) := g\Big(\mathbf{q}^*(G, \mathbf{v}(\mathbf{x}^*), F_\epsilon(\cdot), \beta), \mathbf{x}^*\Big) - g\Big(\mathbf{q}^*(G, \mathbf{v}(\mathbf{x}), F_\epsilon(\cdot), \beta), \mathbf{x}\Big), \tag{18}$$

where $\mathbf{x}^*$ is the optimal solution derived from original problem (4).

6.1. Influence Maximization

In the IM problem, the goal is to select up to K seed users to adopt a service initially, so as to maximize the long-term expected total adoptions of the entire network. For example, a service provider may target key influencers on social media as the initial adopters to stimulate broader adoption. We assume that the adoptions of seed users are irreversible, contrasting the typical nonprogressive setting. This assumption serves two main purposes. First, it reflects the practical scenario where the influence of initial seed users is long-lasting, as is often observed in real-world scenarios. Second, as shown in Proposition 1, altering only the initial adoption states does not affect the long-term equilibrium. By making seed user adoptions irreversible, we effectively modify the system's limiting behavior. This assumption can also be interpreted as increasing the intrinsic values of the seed users to sufficiently high levels, ensuring that they will always adopt the service.

Given diffusion instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, the original IM problem can be formulated as

$$\underset{S\subseteq V:|S|=K}{\text{maximize}} \quad \sum_{i\in V} q_i^* = \lim_{t\to\infty} \sum_{i\in V} \mathbb{E}[Y_i(t)] \tag{19a}$$

$$\text{subject to} \quad Y_i(t) = \begin{cases} 1 & \forall i \in S, t \geq 1, \\ \mathbb{1}\left\{v_i + \beta \frac{\sum_{j\in\mathcal{N}_i} Y_i(t-1)}{d_i} + \epsilon_i(t) \geq 0\right\} & \forall i \in V\setminus S, t \geq 1. \end{cases} \tag{19b}$$

The objective (19a) is the limiting total expected adoptions and (19b) describes the MC that determines $\mathbf{q}^*$ with initialization $\mathbf{Y}(0) = \mathbf{1}$ and $Y_i(t) = 1$ for all $i \in S$ and $t \geq 1$.

Employing the FPA scheme, the approximate IM problem can be formalized as follows:

$$\underset{\boldsymbol{\mu}, S\subseteq V:|S|=K}{\text{maximize}} \quad \sum_{i\in V} \mu_i \tag{20a}$$

$$\text{subject to} \quad \mu_i = 1, \qquad \forall i \in S, \tag{20b}$$

$$\mu_i = 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \mu_j}{d_i}\right), \ \forall i \in V\setminus S. \tag{20c}$$

For ease of formulation, we use $\boldsymbol{\mu}$ as an explicit decision variable and a set of equality constraints specifies the FPA scheme which uniquely determines $\boldsymbol{\mu}^*(S)$ for any given $S \subseteq V$. We next derive a regret bound for the optimal solution S^{FPA} to (20) compared to the optimal solution to (19), S^*.

PROPOSITION 3 **(Regret Bound for the IM Problem)**. *Under Assumptions 1 and 2, for any IM instance* $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, $\text{Regret}(S^{\text{FPA}}) \leq 2C_\rho\sqrt{n\|\mathcal{C}(G;\rho)\|_1}$.

The approximation error of the FPA scheme directly translates into a decision error. All of our previous findings, including the refined bounds discussed in Section 4, extend to the approximate IM problem. For example, following (7), the worst-case regret bound in Proposition 3 can also be adjusted to an order of $\mathcal{O}(n/\sqrt{d_{\min}})$, which is sublinear in n when $d_{\min}$ increases with rate $\Omega(1)$.

Although (20) provides a good approximation to the IM problem (19), solving it remains a challenging task. Under a mild technical condition stated below, we prove $g(\boldsymbol{\mu}^*(S), S)$ is submodular with regard to the seed set S.

ASSUMPTION 4 **(Restricted Convexity of the CDF)**. *The random noise CDF $F_\epsilon(\cdot)$ is convex on the domain $[-\max_{i\in V} v_i - \beta, -\min_{i\in V} v_i]$.*

Assumption 4 covers a wide range of commonly studied cases. For example, the nonprogressive LT model, a special case of our model, naturally satisfies this assumption. Detailed discussions and additional examples supporting this assumption can be found in Appendix F.1. Under Assumption 4, the following theorem shows the submodularity of $g(\boldsymbol{\mu}^*(S), S)$.

THEOREM 4 **(Submodularity of Approximate IM Problem)**. *Under Assumptions 1, 2 and 4, $g(\boldsymbol{\mu}^*(S), S)$ is a submodular set function with regard to seed set S.*

As an immediate consequence of Theorem 4, the well-known greedy algorithm (e.g., see Nemhauser et al. 1978), which recursively adds the node with the largest marginal increase in total adoptions (i.e., adding node i that maximizes $\boldsymbol{\mu}^*(S\cup\{i\}) - \boldsymbol{\mu}^*(S)$) is applicable for solving the approximate IM problem. For instances that satisfy Assumption 4, the greedy algorithm provides a $(1-1/e)$-approximation solution to problem (20). Together with Proposition 3, the simple greedy approach provides a high-quality solution to the original IM problem (19).

In summary, our approximate IM formulation offers several advantages. It allows us to establish a clear condition, i.e., Assumption 4, ensuring submodularity, whereas verifying a similar condition for the original problem is challenging. Furthermore, the greedy algorithm is more efficient when applied to the approximate IM problem, as the FPA scheme speeds up the calculation of limiting adoption probabilities, which must be computed $\mathcal{O}(nK)$ times. To validate our findings, we conduct extensive numerical experiments, with details provided in Appendix F.2. Our results show that, regardless of Assumption 4, the greedy algorithm for the approximate IM problem achieves near-optimal solutions. Moreover, it outperforms many heuristics and performs nearly as well as the greedy solution for the original problem while running much faster.

6.2. Optimal Pricing on A Social Network

Network effects play an important role in shaping customer preferences for products or services, which has driven a growing body of literature that integrates network effects into revenue management problems (Du et al. 2016, 2018, Wang and Wang 2017, Chen and Shi 2023, Chen and Chen 2021, Gopalakrishnan et al. 2022). Our model naturally connects to this literature, which often employs axiomatic or game-theoretic frameworks that approximate the limiting behavior of customer purchasing decisions in dynamic, stochastic environments. In this section, we first follow the literature by discussing the general static pricing problem under limiting behaviors. We then extend this framework to dynamic pricing strategies for transient periods, assuming the platform can exercise perfect price discrimination. This is made possible by our observation discussed in Section 3.3 that the FPA scheme not only approximates the limiting behavior but also effectively captures the entire diffusion trajectory in transient periods.

6.2.1. Static pricing problem under limiting behavior. We consider a firm using pricing as an operational lever to steer consumers' adoption decisions. The adoption utility for user i at time t is given by $u_i(t) = v_i - \gamma p_i + \beta \cdot \frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i} + \epsilon_i(t)$, where p_i is the price offered to user i and γ denotes the price sensitivity. We allow for different prices to be offered to different consumers, as many platforms are capable of targeted price discrimination. Suppose the platform can set at most m distinct prices, represented by $\mathbf{p} \in \mathbb{R}^m$. We define a transformation matrix $\mathbf{W} \in \mathbb{R}^{n\times m}$, where $W_{ik} = 1$ if consumer i is assigned to the k-th price, and $W_i = 0$ otherwise. When $m = n$ and $\mathbf{W} = \mathbf{I}_n$, customers face idiosyncratic prices. When $m = 1$ and $\mathbf{W} = \mathbf{e}_n$, customers face a homogeneous price. Intermediate forms of price discrimination are also possible, such as pricing based on a customer's network connectivity. We assume that the transformation matrix $\mathbf{W}$ is known to the platform, meaning that the platform has pre-determined m customer segments for pricing purposes. The ultimate objective is to identify an optimal price vector that maximizes total profit.

Given diffusion instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, the original pricing problem can be formulated as

$$\underset{\mathbf{p}}{\text{maximize}} \quad \sum_{i\in V}\left(\sum_{k=1}^{m} W_{ik}p_k\right)\cdot q_i^* = \sum_{i\in V}\left(\sum_{k=1}^{m} W_{ik}p_k\right)\cdot \lim_{t\to\infty}\mathbb{E}[Y_i(t)] \tag{21a}$$

$$\text{subject to} \quad Y_i(t) = \mathbb{1}\left\{v_i - \gamma\sum_{k=1}^{m} W_{ik}p_k + \beta\frac{\sum_{j\in\mathcal{N}_i} Y_i(t-1)}{d_i} + \epsilon_i(t) \geq 0\right\} \quad \forall i \in V, t \geq 1. \tag{21b}$$

The objective (21a) is the total profit and (21b) describes the MC that determines $\mathbf{q}^*$ with given price $\mathbf{p}$. Employing the FPA scheme, the approximate problem can be formally stated as:

$$\underset{\boldsymbol{\mu},\mathbf{p}}{\text{maximize}} \quad \boldsymbol{\mu}^\top \mathbf{W}\mathbf{p} \tag{22a}$$

$$\text{subject to} \quad \mu_i = 1 - F_\epsilon\left(-v_i + \gamma\sum_{k=1}^{m} W_{ik}p_k - \beta\frac{\sum_{j\in\mathcal{N}_i}\mu_j}{d_i}\right), \ \forall i \in V. \tag{22b}$$

We treat $\boldsymbol{\mu}$ as an explicit decision variable and use (22b) to link $\mathbf{p}$ and $\boldsymbol{\mu}$. The approximate problem (22) is generally nonconvex due to constraint (22b), making it challenging to solve. Similar to the IM problem, we first establish the regret bound for the optimal solution to (22), denoted by $\mathbf{p}^{\text{FPA}}$.

PROPOSITION 4 **(Regret Bound for the Pricing Problem)**. *Under Assumptions 1 and 2, for any pricing instance* $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, $\text{Regret}(\mathbf{p}^{\text{FPA}}) \leq p_{\max} C_\rho \sqrt{n\|\mathcal{C}(G;\rho)\|_1}$, *where* $p_{\max} := \max\left\{\|\mathbf{p}^*\|_\infty, \|\mathbf{p}^{\text{FPA}}\|_\infty\right\}$.

Proposition 4 establishes a regret bound similar to Proposition 3, except that the bound depends on $p_{\max}$, which can be seen as the maximal derivative of objective function with regard to the adoption probability. In practice, the platform usually has a natural upper bound for the prices (e.g., the price under which no agent will make a purchase, regardless of their neighbors' adoption

decisions). Therefore, $p_{\max}$ can be bounded by a constant. Hence, similar to Proposition 3, the bound in Proposition 3 also highlights a sublinear dependency on n and the network structure.

The distribution of random noise affects both the formulation and hardness of the problem. For certain distributions, such as normal distribution, the problem becomes more complex. Hereafter, we focus on the standard logistic distribution, i.e., $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0,1)$. This choice aligns the formulation with existing existing revenue management literature (Li and Huh 2011, Gallego and Wang 2014, Golrezaei et al. 2020, Chen and Shi 2023). A well-established technique for analyzing such problems is to transform them into optimization problems in demand space. Motivated by this approach, we consider the problem in both the adoption probability and price spaces.

Profit maximization in the adoption probability space. When considering the adoption probability space, the pricing problem becomes less challenging for certain cases. Under perfect price discrimination ($m = n$, $\mathbf{W} = \mathbf{I}_n$) where the platform can provide an idiosyncratic price/subsidy to each consumer and without additional constraints, the problem can be expressed as follows[6]:

$$\underset{\boldsymbol{\mu},\mathbf{p}}{\text{maximize}} \quad \boldsymbol{\mu}^\top \mathbf{p} \tag{23a}$$

$$\text{subject to} \quad \mu_i = 1 - \frac{1}{1+\exp\{v_i + \beta \sum_{j\in\mathcal{N}_i} \mu_j/d_i - \gamma p_i\}}, \ \forall i \in V. \tag{23b}$$

Cancelling out $\mathbf{p}$, we can reformulate the problem in the adoption probability space as:

$$\underset{\boldsymbol{\mu}}{\text{maximize}} \quad \sum_{i\in V} \frac{1}{\gamma} \left(v_i + \beta \sum_{j\in\mathcal{N}_i} \frac{\mu_j}{d_i} + \ln \frac{1-\mu_i}{\mu_i} \right) \mu_i \tag{24a}$$

$$\text{subject to} \quad 0 \le \mu_i \le 1, \ \forall i \in V. \tag{24b}$$

When $\beta = 0$, the network effect term has no impact, and the problem reduces to the classical pricing problem with a concave objective. We show that this property is preserved when β is small.

THEOREM 5 **(Concavity of Static Pricing Objective)**. *The objective of the static pricing problem* (24) *is concave in* $\boldsymbol{\mu}$ *when* $0 \le \beta \le 3.375$.

Theorem 5 states that when $0 \le \beta \le 3.375$, problem (24) becomes a convex optimization problem, and the optimal solution $\boldsymbol{\mu}^*$ can be achieved by standard optimization techniques (i.e.. gradient methods). Once $\boldsymbol{\mu}^*$ is obtained, the optimal prices can be recovered. Furthermore, we remark that both Theorem 5 and Assumption 2 require the network effect intensity to be relatively small, which is a condition consistently made in the related literature.

Profit maximization in the price space. In a more general setting where perfect price discrimination is not feasible, the pricing problem cannot be reformulated in the adoption probability

[6] Negative prices are allowed, meaning the platform can subsidize certain users, particularly those with a large influence on the network. In such cases, the platform incurs losses for these customers to generate a larger overall profit, which is a common practice.

space. Thus, we need to study profit maximization directly in the price space. Particularly, we represent the adoption probability as an implicit function of price, $\boldsymbol{\mu}(\mathbf{p})$, and write the profit function as $\Pi(\mathbf{p})$. We can then derive the gradient of $\Pi(\mathbf{p})$ as follows:

$$\frac{d\Pi(\mathbf{p})}{d\mathbf{p}} = \frac{d\boldsymbol{\mu}(\mathbf{p})}{d\mathbf{p}} \cdot \mathbf{W} \cdot \mathbf{p} + \mathbf{W}^\top \cdot \boldsymbol{\mu}(\mathbf{p}), \tag{25}$$

where the gradient of $\boldsymbol{\mu}(\mathbf{p})$ is not explicitly given. To obtain this, we apply the implicit function theorem to (23b) (i.e., $\boldsymbol{\mu}(\mathbf{p}) = \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))$; see Appendix F.3 for details) and rewrite (25) as

$$\frac{d\Pi(\mathbf{p})}{d\mathbf{p}} = \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \mathbf{p}} \cdot \left(\mathbf{I} - \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})} \right)^{-1} \cdot \mathbf{W} \cdot \mathbf{p} + \mathbf{W}^\top \cdot \boldsymbol{\mu}(\mathbf{p}). \tag{26}$$

We can now apply standard gradient descent techniques to find the near-optimal solution. As a final remark, the profit maximization in the price space as well as the gradient-based approach also hold for other noise distributions or with more complicated price constraints (e.g., the box constraints).

6.2.2. Dynamic pricing problem under transient behavior. Solving the general dynamic pricing problem is challenging, but the special case of perfect price discrimination can be addressed through reformulation in the adoption probability space, similar to the static problem (24). We can then formulate the (approximate) dynamic pricing problem over T time periods as follows:

$$\begin{aligned} \underset{\boldsymbol{\mu}, \mathbf{p}}{\text{maximize}} \quad & \sum_{t=1}^{T} \boldsymbol{\mu}(t)^\top \mathbf{p}(t) \\ \text{subject to} \quad & \mu_i(t) = 1 - \frac{1}{1 + \exp\{v_i + \beta \sum_{j \in \mathcal{N}_i} \mu_j(t-1)/d_i - \gamma p_i(t)\}}, \ \forall i \in V, t \geq 1, \end{aligned}$$

where we initialize the adoption probability as $\boldsymbol{\mu}(0) = \mathbf{0}$. Similar to the static problem, we can eliminate $\mathbf{p}$ and reformulate the dynamic problem in the adoption probability space as:

$$\underset{\boldsymbol{\mu}}{\text{maximize}} \quad \sum_{t=1}^{T} \sum_{i \in V} \frac{1}{\gamma} \left(v_i + \beta \sum_{j \in \mathcal{N}_i} \frac{\mu_j(t-1)}{d_i} + \ln \frac{1 - \mu_i(t)}{\mu_i(t)} \right) \mu_i(t) \tag{27a}$$

$$\text{subject to} \quad 0 \leq \mu_i(t) \leq 1, \ \forall i \in V, t \geq 1. \tag{27b}$$

In Theorem 6 below, we show that the concavity of objective function is preserved under the same conditions for the dynamic pricing problem as the static case. Therefore, the approximate problem can be readily solved. Further, using the arguments in Section 3.3, one can also derive a bounded on the regret of this solution in a similar fashion to Proposition 4. We skip the details for brevity.

THEOREM 6 **(Concavity of Dynamic Pricing Objective)**. *The objective of the dynamic pricing problem* (27) *is concave in* $\boldsymbol{\mu}$ *when* $0 \leq \beta \leq 3.375$.

Finally, we conduct numerical experiments on the pricing problem. We begin by implementing gradient-based algorithms to solve the static pricing problem under two extreme scenarios: (i) perfect price discrimination, where each consumer is offered a personalized price, and (ii) uniform pricing, where all consumers receive the same price. In both cases, we show that near-optimal solutions can be achieved. Additionally, for the perfect price discrimination case, we extend our experiments to the dynamic pricing problem and compare the results with those from the static pricing problem. Our findings reveal that a myopic pricing strategy is suboptimal. The optimal dynamic pricing strategy uses price as a tool to quickly guide consumer behavior toward the limiting behavior, followed by a period of stable pricing close to the static optimal price. Notably, during the transient period, the optimal price tends to increase, indicating that a lower price is initially used to attract more consumers before stabilizing at a higher price. For a more detailed discussion of the implementation and results, interested readers can refer to Appendices F.3 and F.4.

7. Conclusion

In this study, we focus on nonprogressive diffusion on a social network, where agents can withdraw their previous decisions in accordance with a change in the social environment. We tide over the issues of the lack of a general modeling framework and efficient algorithms in the previous studies. Specifically, we base on a general nonprogressive diffusion model that is agent-based, considers the local network effect, and can be adapted to many utility models. We propose, with a provable performance guarantee, a fixed-point approximation scheme that can accurately and efficiently approximate the adoption behavior for all agents and validate the results through extensive experiments. We provide order-optimal bounds for the approximation error and conduct a thorough analysis of its dependency with network structure. Finally, we investigate the conventional optimization problems based on the fixed-point approximation.

We also view one of our contributions as proposing a novel approach to studying the user behavior within networks in a stochastic setting. In particular, there are several directions for future research, in which our method seems readily extendable. First, the adoptions may not change in each period but last for several periods in practice (e.g., a user needs to subscribe to Netflix for at least one month). It would be interesting to investigate how we can represent the limiting behavior in this scenario. Second, this work only considers a binary-choice case where each agent only decides to adopt or not. It is worth investigating whether similar results can be extended to a multiple-choice case (e.g., not to subscribe, to subscribe to a normal membership, or to subscribe to a premium membership). Finally, the local network effect is captured by the average adoption of the in-neighbors in our model. It is promising to consider the weighted average of in-neighbor adoptions where the network effect is asymmetric.

References

Acemoğlu, D., Como, G., Fagnani, F., and Ozdaglar, A. (2013). Opinion fluctuations and disagreement in social networks. Mathematics of Operations Research, 38(1):1–27.

Acemoglu, D., Dahleh, M. A., Lobel, I., and Ozdaglar, A. (2011). Bayesian learning in social networks. The Review of Economic Studies, 78(4):1201–1236.

Afèche, P., Liu, Z., and Maglaras, C. (2023). Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance. Manufacturing & Service Operations Management, 25(5):1890–1908.

Agrawal, S., Yin, S., and Zeevi, A. (2021). Dynamic pricing and learning under the bass model. In Proceedings of the 22nd ACM Conference on Economics and Computation, pages 2–3.

Allon, G., Drakopoulos, K., and Manshadi, V. (2019). Information inundation on platforms and implications. In Proceedings of the 2019 ACM Conference on Economics and Computation, pages 555–556.

Anari, N., Ehsani, S., Ghodsi, M., Haghpanah, N., Immorlica, N., Mahini, H., and Mirrokni, V. S. (2010). Equilibrium pricing with positive externalities. In International Workshop on Internet and Network Economics, pages 424–431. Springer.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. wanted: The key player. Econometrica, 74(5):1403–1417.

Bapna, R. and Umyarov, A. (2015). Do your online friends make you pay? a randomized field experiment on peer influence in online social networks. Management Science, 61(8):1902–1920.

Baron, O., Hu, M., and Malekian, A. (2022). Revenue volatility under uncertain network effects. Operations Research.

Bass, F. M. (1969). A new product growth for model consumer durables. Management science, 15(5):215–227.

Benaïm, M. and Weibull, J. W. (2003). Deterministic approximation of stochastic evolution in games. Econometrica, 71(3):873–903.

Bonacich, P. (1987). Power and centrality: A family of measures. American journal of sociology, 92(5):1170–1182.

Boykov, Y., Veksler, O., and Zabih, R. (1998). Markov random fields with efficient approximations. In Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231), pages 648–655. IEEE.

Candogan, O., Bimpikis, K., and Ozdaglar, A. (2012). Optimal pricing in networks with externalities. Operations Research, 60(4):883–905.

Chandrasekhar, A. G., Larreguy, H., and Xandri, J. P. (2020). Testing models of social learning on networks: Evidence from two experiments. Econometrica, 88(1):1–32.

Chen, N. and Chen, Y.-J. (2021). Duopoly competition with network effects in discrete choice models. Operations Research, 69(2):545–559.

Chen, W., Wang, Y., and Yang, S. (2009). Efficient influence maximization in social networks. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 199–208.

Chen, W., Yuan, Y., and Zhang, L. (2010). Scalable influence maximization in social networks under the linear threshold model. In 2010 IEEE international conference on data mining, pages 88–97. IEEE.

Chen, X., van der Lans, R., and Trusov, M. (2021). Efficient estimation of network games of incomplete information: Application to large online social networks. Management Science, 67(12):7575–7598.

Chen, Y. and Shi, C. (2023). Network revenue management with online inverse batch gradient descent method. Production and Operations Management.

Datareportal (2022). Digital 2022: Global overview report. https://datareportal.com/reports/digital-2022-global-overview-report. Accessed: 2022-06-27.

Drakopoulos, K. and Zheng, F. (2017). Network effects in contagion processes: Identification and control. Columbia Business School Research Paper, (18-8).

Du, C., Cooper, W. L., and Wang, Z. (2016). Optimal pricing for a multinomial logit choice model with network effects. Operations Research, 64(2):441–455.

Du, C., Cooper, W. L., and Wang, Z. (2018). Optimal worst-case pricing for a logit demand model with network effects. Operations Research Letters, 46(3):345–351.

Erdős, P., Rényi, A., et al. (1960). On the evolution of random graphs. Publ. math. inst. hung. acad. sci, 5(1):17–60.

Esary, J. D., Proschan, F., and Walkup, D. W. (1967). Association of random variables, with applications. The Annals of Mathematical Statistics, 38(5):1466–1474.

Feng, Q., Li, C., Lu, M., and Shanthikumar, J. G. (2022). Implementing environmental and social responsibility programs in supply networks through multiunit bilateral negotiation. Management Science, 68(4):2579–2599.

Gallego, G. and Wang, R. (2014). Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. Operations Research, 62(2):450–461.

Göbel, F. and Jagers, A. (1974). Random walks on graphs. Stochastic processes and their applications, 2(4):311–336.

Goldenberg, J., Libai, B., and Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing letters, 12(3):211–223.

Goldstein, L. and Wiroonsri, N. (2018). Stein's method for positively associated random variables with applications to the ising and voter models, bond percolation, and contact process.

Golrezaei, N., Jaillet, P., and Liang, J. C. N. (2020). No-regret learning in price competitions under consumer reference effects. Advances in Neural Information Processing Systems, 33:21416–21427.

Gopalakrishnan, M., Zhang, H., and Zhang, Z. (2022). Multiproduct pricing under the multinomial logit model with local network effects. Decision Sciences.

Granovetter, M. (1978). Threshold models of collective behavior. American journal of sociology, 83(6):1420–1443.

Horst, U. and Scheinkman, J. A. (2006). Equilibria in systems of social interactions. Journal of Economic Theory, 130(1):44–77.

Hu, M., Wang, Z., and Feng, Y. (2020). Information disclosure and pricing policies for sales of network goods. Operations Research, 68(4):1162–1177.

Huang, J., Mani, A., and Wang, Z. (2022). The value of price discrimination in large social networks. Management Science, 68(6):4454–4477.

Ising, E. (1924). Beitrag zur theorie des ferro-und paramagnetismus. PhD thesis, Grefe & Tiedemann Hamburg, Germany.

Jackson, M. O. (2010). Social and economic networks. Princeton university press.

Jackson, M. O., Lin, Z., and Yu, N. N. (2020). Adjusting for peer-influence in propensity scoring when estimating treatment effects. Available at SSRN 3522256.

Jadbabaie, A., Molavi, P., Sandroni, A., and Tahbaz-Salehi, A. (2012). Non-bayesian social learning. Games and Economic Behavior, 76(1):210–225.

Janson, S., Knuth, D. E., Łuczak, T., and Pittel, B. (1993). The birth of the giant component. Random Structures & Algorithms, 4(3):233–358.

Kempe, D., Kleinberg, J., and Tardos, É. (2003). Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 137–146.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character, 115(772):700–721.

Lee, L.-f., Li, J., and Lin, X. (2014). Binary choice models with social network under heterogeneous rational expectations. Review of Economics and Statistics, 96(3):402–417.

Li, H. (2020). Optimal pricing under diffusion-choice models. Operations Research, 68(1):115–133.

Li, H. and Huh, W. T. (2011). Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications. Manufacturing & Service Operations Management, 13(4):549–563.

Li, Y., Fan, J., Wang, Y., and Tan, K.-L. (2018). Influence maximization on social graphs: A survey. IEEE Transactions on Knowledge and Data Engineering, 30(10):1852–1872.

Lin, Y., Wang, M., Shen, Z.-J. M., Zhang, H., and Zhang, R. P. (2021). Content promotion for online content platforms with network diffusion effect. Available at SSRN 3863104.

Lu, Y., Jerath, K., and Singh, P. V. (2013). The emergence of opinion leaders in a networked online community: A dyadic model with time dynamics and a heuristic for fast estimation. Management Science, 59(8):1783–1799.

Ma, L., Krishnan, R., and Montgomery, A. L. (2015). Latent homophily or social influence? an empirical analysis of purchase within a social network. Management Science, 61(2):454–473.

Mislove, A. E. (2009). Online social networks: measurement, analysis, and applications to distributed information systems. Rice University.

Molloy, M. and Reed, B. (1995). A critical point for random graphs with a given degree sequence. Random structures & algorithms, 6(2-3):161–180.

Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. (1978). An analysis of approximations for maximizing submodular set functions—i. Mathematical programming, 14(1):265–294.

Newman, C. M. (1980). Normal fluctuations and the fkg inequalities. Communications in Mathematical Physics, 74(2):119–128.

Newman, M. E., Strogatz, S. H., and Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. Physical review E, 64(2):026118.

Nosrat, F., Cooper, W. L., and Wang, Z. (2021). Pricing for a product with network effects and mixed logit demand. Naval Research Logistics (NRL), 68(2):159–182.

Rheinboldt, W. C. (1998). Methods for solving systems of nonlinear equations. SIAM.

Rossi, R. A. and Ahmed, N. K. (2015). The network data repository with interactive graph analytics and visualization. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

Sadler, E. (2020). Diffusion games. American Economic Review, 110(1):225–70.

Schelling, T. C. (1978). Micromotives and macrobehavior. Technical report.

Shen, S., You, M., and Ma, Y. (2017). Single-commodity stochastic network design under demand and topological uncertainties with insufficient data. Naval Research Logistics (NRL), 64(2):154–173.

Shriver, S. K., Nair, H. S., and Hofstetter, R. (2013). Social ties and user-generated content: Evidence from an online social network. Management Science, 59(6):1425–1443.

Song, J.-S. and Zipkin, P. (2009). Inventories with multiple supply sources and networks of queues with overflow bypasses. Management Science, 55(3):362–372.

Van Mieghem, P., Omic, J., and Kooij, R. (2008). Virus spread in networks. IEEE/ACM Transactions On Networking, 17(1):1–14.

Wang, R. and Wang, Z. (2017). Consumer choice models with endogenous network effects. Management Science, 63(11):3944–3960.

Xie, T. and Wang, Z. (2020). Personalized assortment optimization under consumer choice models with local network effects. Available at SSRN 3788880.

Xu, H. (2018). Social interactions in large networks: A game theoretic approach. International Economic Review, 59(1):257–284.

Yang, N. and Zhang, R. P. (2022). Dynamic pricing and inventory management in the presence of online reviews. Production and Operations Management, 31(8):3180–3197.

Yeomans, J. M. (1992). Statistical mechanics of phase transitions. Clarendon Press.

Zeng, Z., Dai, H., Zhang, D. J., Zhang, H., Zhang, R., Xu, Z., and Shen, Z.-J. M. (2023). The impact of social nudges on user-generated content for social network platforms. Management Science, 69(9):5189–5208.

Online Appendices

Appendix A: Supporting Arguments for Section 2

In this Appendix, we provide the proofs for the theoretical results in Section 2, focusing primarily on the proof of Proposition 1. Before proceeding with that, we present the following lemma and its proof, which will be useful in establishing Proposition 1.

LEMMA 5. *Given an arbitrary agent $i \in V$, at least one of the following values*

$$(a) \prod_{\mathbf{y}\in\{0,1\}^{|V|}} \mathbb{P}(Y_i(t)=1|\mathbf{Y}(t-1)=\mathbf{y}) \text{ or } (b) \prod_{\mathbf{y}\in\{0,1\}^{|V|}} \mathbb{P}(Y_i(t)=0|\mathbf{Y}(t-1)=\mathbf{y}),$$

are positive.

Proof of Lemma 5: When random noise $\epsilon_i(t)$ is not bounded on either side, it is obvious that the statement holds. In the following, we only consider the situation when $\epsilon_i(t)$ is with support on some bounded interval $[\underline{\epsilon}, \overline{\epsilon}]$. We will show by contradiction.

If (a) is not positive, then there exists $\mathbf{y} \in \{0,1\}^{|V|}$ such that $\mathbb{P}(Y_i(t)=1|\mathbf{Y}(t-1)=\mathbf{y})=0$, which means $F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{d_i}\right)=1$. Hence, we have

$$-v_i - \inf_{\mathbf{y}\in\{0,1\}^{|V|}} \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{d_i} = -v_i \geq \overline{\epsilon}.$$

Consequently, we can derive the following inequality

$$-v_i - \sup_{\mathbf{y}\in\{0,1\}^{|V|}} \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{d_i} = -v_i - \beta > -v_i - \frac{1}{L} \geq -v_i - (\overline{\epsilon}-\underline{\epsilon}) \geq \underline{\epsilon},$$

where the first inequality follows from Assumption 2, the second inequality follows Assumption 1, that is, $|F_\epsilon(\overline{\epsilon}) - F_\epsilon(\underline{\epsilon})| \leq L|\overline{\epsilon}-\underline{\epsilon}|$. As a direct result, for any $\mathbf{y}\in\{0,1\}^{|V|}$, $F_\epsilon(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{d_i}) \geq F_\epsilon(\underline{\epsilon}) > 0$ and thus the value of (b) is positive.

In conclusion, for each agent $i \in V$, at least one of (a) and (b) have positive value. □

Proof of Proposition 1: Our goal is to show that the MC has only one recurrent communication class and is aperiodic, regardless of whether it is irreducible.

We first prove that there is only one recurrent communication class. To do this, we construct the vector $\mathbf{y}'$ as follows

$$y_i' = \begin{cases} 0 & \text{when } \prod_{\mathbf{y}\in\{0,1\}^{|V|}} \mathbb{P}(Y_i(t)=1|\mathbf{Y}(t-1)=\mathbf{y})=0, \\ 1 & \text{o.w.} \end{cases}$$

Following Lemma 5, for any index i such that $y_i'=0$, we have $\mathbb{P}(Y_i(t)=0|\mathbf{Y}(t-1)=\mathbf{y})>0$. For indices i such that $y_i'=1$, we have $\mathbb{P}(Y_i(t)=1|\mathbf{Y}(t-1)=\mathbf{y})>0$. Consequently, for all $i\in V$ and $\mathbf{y}\in\{0,1\}^{|V|}$, we have $\mathbb{P}(Y_i(t)=y_i'|\mathbf{Y}(t-1)=\mathbf{y})>0$.

For any arbitrary $\mathbf{y}\in\{0,1\}^{|V|}$, the transition probability between $\mathbf{y}$ and $\mathbf{y}'$ can be written as:

$$P(\mathbf{y},\mathbf{y}') = \prod_{i\in V} \mathbb{P}(Y_i(t)=y_i'|\mathbf{Y}(t-1)=\mathbf{y}),$$

which, based on our analysis, will be positive. This implies that all states in the MC communicate with the state $\mathbf{y}'$. Therefore, the set of states that $\mathbf{y}'$ communicates with forms a recurrent communication class, while all other states are in transient classes.

Furthermore, since $P(\mathbf{y}', \mathbf{y}') > 0$ also holds, state $\mathbf{y}'$ has period 1 which implies that the recurrent communication class of the MC is aperiodic.

In conclusion, this MC has a limiting distribution $\boldsymbol{\pi}$ that satisfies $\boldsymbol{\pi} = \boldsymbol{\pi} P$ and the limiting adoption probability of each agent is a linear transformation of $\boldsymbol{\pi}$ that follows (2). □

Appendix B: Supporting Arguments for Section 3

In this Appendix, we provide the proofs for the theoretical results in Section 3. The main objective of Section 3 is to prove our key result, Theorem 1. Due to the length of the proof, we decompose it into three parts for clarity and ease of reading. We begin by establishing the existence and uniqueness of the FPA solution in Appendix B.1. Next, we analyze the covariance across agents, which is the most complex part, in Appendix B.2. Finally, we conclude the proof for Theorem 1 and its associated corollary in Appendix B.3.

B.1. Supporting Arguments for Proposition 2

Proof of Proposition 2: We first show the property (i) and then prove properties (ii) and (iii) by showing that $\mathbf{h}(\cdot)$ is a contraction mapping.

Proof of (i): When $\boldsymbol{a} \leq \mathbf{b}$, we have $\sum_{j\in\mathcal{N}_i} a_j \leq \sum_{j\in\mathcal{N}_i} b_j$ for all $i \in V$. Since CDF $F_\epsilon(\cdot)$ is monotonically nondecreasing, if $\boldsymbol{a} \leq \mathbf{b}$,

$$1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} a_j}{d_i}\right) \leq 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} b_j}{d_i}\right),$$

for all $i \in V$, which implies $\mathbf{h}(\boldsymbol{a}) \leq \mathbf{h}(\mathbf{b})$.

Proof of (ii) and (iii): It is trivial that $\mathbf{h}(\cdot)$ maps $\mathbb{R}^n$ to itself. Consider the Jacobian matrix of $\mathbf{h}(\boldsymbol{\mu})$, for an arbitrary $\boldsymbol{\mu} \in \mathbb{R}^n$, we have

$$\frac{\partial h(\boldsymbol{\mu})_i}{\partial \mu_j} = \begin{cases} 0, & j \notin \mathcal{N}_i \\ \dfrac{\beta}{d_i} \dfrac{\partial F_\epsilon\left(-v_i - \beta \dfrac{\sum_{j'\in\mathcal{N}_i} \mu_{j'}}{d_i}\right)}{\partial\left(-v_i - \beta \dfrac{\sum_{j'\in\mathcal{N}_i} \mu_{j'}}{d_i}\right)}, & j \in \mathcal{N}_i. \end{cases}$$

By Assumption 1, we can have $\left|\frac{\partial h(\boldsymbol{\mu})_i}{\partial \mu_j}\right| \leq \frac{\beta L}{d_i}$ for all $j \in \mathcal{N}_i$. Therefore, the ∞-norm of $\frac{d\mathbf{h}(\boldsymbol{\mu})}{d\boldsymbol{\mu}}$ can be upper bounded as

$$\left\|\frac{d\mathbf{h}(\boldsymbol{\mu})}{d\boldsymbol{\mu}}\right\|_\infty = \max_{i\in V} \sum_{j\in V} \left|\frac{\partial h(\boldsymbol{\mu})_i}{\partial \mu_j}\right| \leq \max_{i\in V} d_i \frac{\beta L}{d_i} = \beta L < 1$$

where the last inequality follows from Assumption 2.

Thus, for all $\boldsymbol{\mu} \in \mathbb{R}^n$, we have $\left\|\frac{d\mathbf{h}(\boldsymbol{\mu})}{d\boldsymbol{\mu}}\right\|_\infty < 1$. It then implies that $\mathbf{h}(\boldsymbol{\mu})$ is a contraction mapping. By the contraction mapping theorem, we conclude the proof. □

B.2. Supporting Arguments for Covariance Analysis

Before we present the proof for Lemma 1, we would like to first present a series of auxiliary definitions and lemmas. To begin with, we define the concept of positive quadrant dependent and positive association in Definitions 1 and 2, respectively.

Definition 1 (Positive Quadrant Dependent, Newman (1980)). A pair of random variables (X,Y) is positively quadrant dependent (PQD) if

$$\mathbb{P}(X \le a, Y \le b) - \mathbb{P}(X \le a)\mathbb{P}(Y \le b) \ge 0, \ \forall\ a,b \in \mathbb{R}.$$

Definition 2 (Positive Association, Esary et al. (1967)). A random vector $\mathbf{X} = (X_1, \ldots, X_n) \in \mathbb{R}^n$ is said to be positively associated (PA) whenever

$$\mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X})) \ge 0,$$

for all coordinate-wise nondecreasing functions f and g on $\mathbb{R}^n$ such that $f(\mathbf{X})$ and $g(\mathbf{X})$ both possess finite second moments.

We then present several lemmas on the properties of variables that are PQD and PA, which paves the way for our covariance analysis.

Lemma 6 **(Esary et al. (1967), P2)**. *If two sets of PA random variables are independent of one another, then their union is a set of PA random variables.*

Proof of Lemma 6: Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, X_2, \ldots, Y_m)$ be PA random vectors, and suppose $\mathbf{X}$ and $\mathbf{Y}$ are independent of each other. Let f, g be any piecewise nondecreasing functions with regard to $(\mathbf{X}, \mathbf{Y})$. Using the law of total covariance, we can decompose the covariance as follows:

$$\mathrm{Cov}\left(f(\mathbf{X},\mathbf{Y}), g(\mathbf{X},\mathbf{Y})\right) = \mathbb{E}_{\mathbf{X}}\left[\mathrm{Cov}_{\mathbf{Y}}(f(\mathbf{X},\mathbf{Y}), g(\mathbf{X},\mathbf{Y})|\mathbf{X})\right] + \mathrm{Cov}_{\mathbf{X}}\left(\mathbb{E}_{\mathbf{Y}}[f(\mathbf{X},\mathbf{Y})|\mathbf{X}], \mathbb{E}_{\mathbf{Y}}[g(\mathbf{X},\mathbf{Y})|\mathbf{X}]\right).$$

Then, it is straightforward to verify that both terms are nonnegative.

(i) First term: Since $\mathbf{Y}$ is PA, we have $\mathrm{Cov}_{\mathbf{Y}}(f(\mathbf{x},\mathbf{Y}), g(\mathbf{x},\mathbf{Y})) \ge 0$ for any given $\mathbf{x}$. Thus, by taking the expectation over this covariance, the first term is nonnegative.

(ii) Second term: Since $\mathbf{X}$ and $\mathbf{Y}$ are independent, thus, we have

$$\mathbb{E}_{\mathbf{Y}}[f(\mathbf{X},\mathbf{Y})|\mathbf{X} = \mathbf{x}] = \mathbb{E}_{\mathbf{Y}}[f(\mathbf{x},\mathbf{Y})].$$

For any arbitrary $\mathbf{x} \in \mathbb{R}^n$, $\mathbb{E}_{\mathbf{Y}}[f(\mathbf{x},\mathbf{Y})]$ and $\mathbb{E}_{\mathbf{Y}}[g(\mathbf{x},\mathbf{Y})]$ are nondecreasing functions with regard to $\mathbf{x}$. Since $\mathbf{X}$ is PA random vector, the covariance between these two nondecreasing functions will also be nonnegative.

In conclusion, the union set of these random variables, i.e., $(\mathbf{X},\mathbf{Y})$, is PA. □

Lemma 7 **(Goldstein and Wiroonsri (2018), Lemma 3.3)**. *The pair of random variables $(f(\mathbf{X}), g(\mathbf{X}))$ is PQD whenever random vector $\mathbf{X}$ is PA, and $f(\cdot)$ and $g(\cdot)$ are coordinate-wise nondecreasing functions of $\mathbf{X}$.*

Proof of Lemma 7: In order to prove that $(f(\mathbf{X}), g(\mathbf{X}))$ is PQD, we need to show that

$$\mathbb{P}(f(\mathbf{X}) \le a, g(\mathbf{X}) \le b) - \mathbb{P}(f(\mathbf{X}) \le a)\mathbb{P}(g(\mathbf{X}) \le b) \ge 0, \ \forall\ a, b \in \mathbb{R}.$$

For any given pair of $(a, b) \in \mathbb{R}^2$, let's define the following nondecreasing functions:

$$f_a(x) = \begin{cases} 0 & x \le a, \\ 1 & x > a, \end{cases} \text{ and } g_b(x) = \begin{cases} 0 & x \le b, \\ 1 & x > b. \end{cases}$$

As a consequence, the composite functions $f_a(f(\cdot))$ and $g_b(g(\cdot))$ are also both nondecreasing functions. Since the random vector $\mathbf{X}$ is PA, we have

$$\begin{aligned} 0 &\le \operatorname{Cov}\Big(f_a(f(\mathbf{X})), g_b(g(\mathbf{X}))\Big) = \operatorname{Cov}\Big(1 - f_a(f(\mathbf{X})), 1 - g_b(g(\mathbf{X}))\Big) \\ &= \mathbb{P}\Big(1 - f_a(f(\mathbf{X})) = 1, 1 - g_b(g(\mathbf{X})) = 1\Big) - \mathbb{P}\Big(1 - f_a(f(\mathbf{X})) = 1\Big)\mathbb{P}\Big(1 - g_b(g(\mathbf{X})) = 1\Big) \\ &= \mathbb{P}\Big(f(\mathbf{X}) \le a, g(\mathbf{X}) \le b\Big) - \mathbb{P}\Big(f(\mathbf{X}) \le a\Big)\mathbb{P}\Big(g(\mathbf{X}) \le b\Big), \end{aligned}$$

where the first inequality follows from the definition of PA random variables and the remaining equalities follow from the definition of covariance.

As a consequence, we have proved that $(f(\mathbf{X}), g(\mathbf{X}))$ is PQD. □

LEMMA 8 **(Goldstein and Wiroonsri (2018), Lemma 3.1; Newman (1980), Lemma 3)**. *Let the random variables X, Y have finite second moments and be PQD. Then for any real-valued functions $f(\cdot)$ and $g(\cdot)$ on $\mathbb{R}$ with their derivatives $f'(\cdot)$, $g'(\cdot)$ bounded,*

$$|\operatorname{Cov}(f(X), g(Y))| \le \|f'\|_\infty \|g'\|_\infty \operatorname{Cov}(X, Y),$$

where $\|\cdot\|_\infty$ denotes the sup norm on $\mathbb{R}$.

Proof of Lemma 8: Let $H_{X,Y}(x, y) = \mathbb{P}(X \le x, Y \le y) - \mathbb{P}(X \le x)\mathbb{P}(Y \le y)$.

The covariance of $f(X)$ and $g(Y)$ can be written as

$$\operatorname{Cov}(f(X), g(Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_{X,Y}(x, y) f'(x) g'(y) dx dy.$$

Thus, by the positivity of $H_{X,Y}$ for PQD random variables, we can arrive at

$$|\operatorname{Cov}(f(X), g(Y))| \le \|f'\|_\infty \|g'\|_\infty \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_{X,Y}(x, y) dx dy = \|f'\|_\infty \|g'\|_\infty \operatorname{Cov}(X, Y).$$

□

COROLLARY 4 **(Lower Bound of Covariance)**. *Let the random variables X, Y have finite second moments and be PQD. Then for any real-valued functions $f(\cdot)$ and $g(\cdot)$ on $\mathbb{R}$ with $f'(\cdot)$, $g'(\cdot)$ bounded,*

$$|\operatorname{Cov}(f(X), g(Y))| \ge \|f'\|_{\infty,\min} \|g'\|_{\infty,\min} \operatorname{Cov}(X, Y),$$

where $\|\cdot\|_{\infty,\min}$ denotes the inferior absolute value on the given domain.

The proof to Corollary 4 is similar to that of Lemma 8, mainly relies on the definition of PQD random variables.

Equipped with these properties, we then prove the upper bound for the covariance matrix in Lemma 1.

Proof of Lemma 1: This proof of the upper bound of the covariance matrix involves two critical steps. First, we need to establish that the adoption state vector $\mathbf{Y}(t) = (Y_1(t), Y_2(t), \ldots, Y_n(t))$ is positively associated for any time $t \geq 0$. Second, we iteratively bound the covariance matrix at each time step, leveraging the PA property.

Positively associated variables: We will show this by induction.

Base case: $t = 0$. The random vector $\mathbf{Y}(0)$ is independent, so it is PA.

To show time $t = s+1$: Assume that the random vector $\mathbf{Y}(s)$ is PA.

The random noise $\epsilon_i(s+1)$ is independent across all agents $i \in V$ and therefore, the vector $\boldsymbol{\epsilon}(s+1)$ is PA. Furthermore, vectors $\mathbf{Y}(s)$ and $\boldsymbol{\epsilon}(s+1)$ are independent of each other. By applying Lemma 6, we can show that the union set of random variables $\mathbf{Y}(s)$ and $\boldsymbol{\epsilon}(s+1)$ is also PA.

For any $i \in V$, the random variable

$$Y_i(s+1) = \mathbb{1}\left\{ v_i + \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(s)}{d_i} + \epsilon_i(s+1) \geq 0 \right\},$$

can be considered as a coordinate-wise nondecreasing function of both $\mathbf{Y}(s)$ and $\boldsymbol{\epsilon}(s+1)$. Let $\mathbf{Y}(s+1) := \boldsymbol{\phi}(\mathbf{Y}(s), \boldsymbol{\epsilon}(s+1)) \in \mathbb{R}^N$ denote this transformation. It can be easily shown that the random vector $\mathbf{Y}(s+1)$ is also a PA vector. Specifically, for any nondecreasing functions f and g,

$$\mathrm{Cov}(f(\mathbf{Y}(s+1)), g(\mathbf{Y}(s+1))) = \mathrm{Cov}\left(f\left(\boldsymbol{\phi}(\mathbf{Y}(s), \boldsymbol{\epsilon}(s+1))\right), g\left(\boldsymbol{\phi}(\mathbf{Y}(s), \boldsymbol{\epsilon}(s+1))\right) \right) \geq 0,$$

where the inequality follows from the definition of PA random variables. Since the composite functions $f(\boldsymbol{\phi}(\cdot))$ and $g(\boldsymbol{\phi}(\cdot))$ are coordinate-wise nondecreasing and the union of $\mathbf{Y}(s)$ and $\boldsymbol{\epsilon}(s+1)$ is PA, the covariance is nonnegative.

As a result, we have shown that $\mathbf{Y}(t)$ is PA for all $t \geq 0$. With this result, we can apply Lemma 7 to demonstrate that the network effect terms $\sum_{j \in \mathcal{N}_i} Y_j(t)$ and $\sum_{j' \in \mathcal{N}_{i'}} Y_{j'}(t)$ are PQD for all $i, i' \in V$ and $t \geq 0$.

Upper bound of the covariance matrix:

In the following, we will iteratively upper bound the covariance between any pair of $(Y_i(t), Y_{i'}(t))$ for all $i, i' \in V$ and $t \geq 0$. Using the law of total covariance, we can decompose $\mathrm{Cov}(Y_i(t), Y_{i'}(t))$ into two parts:

$$\begin{aligned}\mathrm{Cov}(Y_i(t), Y_{i'}(t)) \;=\; & \mathbb{E}_{\mathbf{Y}(t-1)}\Big[\mathrm{Cov}_{\boldsymbol{\epsilon}(t)}\left(Y_i(t), Y_{i'}(t) \,\middle|\, \mathbf{Y}(t-1)\right)\Big] \\ & + \mathrm{Cov}_{\mathbf{Y}(t-1)}\Big(\mathbb{E}_{\boldsymbol{\epsilon}(t)}\big[Y_i(t) \;\big|\; \mathbf{Y}(t-1)\big], \mathbb{E}_{\boldsymbol{\epsilon}(t)}\big[Y_{i'}(t) \;\big|\; \mathbf{Y}(t-1)\big]\Big).\end{aligned}$$

The first term $\mathbb{E}_{\mathbf{Y}(t-1)}\left[\mathrm{Cov}_{\boldsymbol{\epsilon}(t)}(Y_i(t), Y_{i'}(t) | \mathbf{Y}(t-1))\right]$ is always 0. The reason is as follows: by applying the law of total conditional covariance again, we have

$$\begin{aligned}& \mathrm{Cov}_{\boldsymbol{\epsilon}(t)}\left(Y_i(t), Y_{i'}(t) \;\middle|\; \mathbf{Y}(t-1)\right) \\ =\; & \mathbb{E}_{\boldsymbol{\epsilon}(t)}\Big[\mathrm{Cov}\left(Y_i(t), Y_{i'}(t) \;\middle|\; \mathbf{Y}(t-1), \boldsymbol{\epsilon}(t)\right)\Big| \; \mathbf{Y}(t-1)\Big] \\ & + \mathrm{Cov}_{\boldsymbol{\epsilon}(t)}\Big(\mathbb{E}\big[Y_i(t) \;\big|\; \mathbf{Y}(t-1), \boldsymbol{\epsilon}(t)\big], \mathbb{E}\big[Y_{i'}(t) \;\big|\; \mathbf{Y}(t-1), \boldsymbol{\epsilon}(t)\big] \;\Big|\; \mathbf{Y}(t-1)\Big).\end{aligned} \tag{28}$$

The former term vanishes because $Y_i(t)$ and $Y_{i'}(t)$ are deterministic when given $\mathbf{Y}(t-1), \boldsymbol{\epsilon}_t$. The latter term is also zero since $\epsilon_i(t)$ and $\epsilon_{i'}(t)$ are independent of each other.

We then show that the second term $\mathrm{Cov}_{\mathbf{Y}(t-1)}\left(\mathbb{E}_{\boldsymbol{\epsilon}(t)}\left[Y_i(t) \mid \mathbf{Y}(t-1)\right],\ \mathbb{E}_{\boldsymbol{\epsilon}(t)}\left[Y_{i'}(t) \mid \mathbf{Y}(t-1)\right]\right)$ can be bounded recursively as

$$\begin{aligned}
&\mathrm{Cov}_{\mathbf{Y}(t-1)}\left(\mathbb{E}_{\boldsymbol{\epsilon}(t)}\left[Y_i(t) \mid \mathbf{Y}(t-1)\right],\ \mathbb{E}_{\boldsymbol{\epsilon}(t)}\left[Y_{i'}(t) \mid \mathbf{Y}(t-1)\right]\right)\\
=\ &\mathrm{Cov}\left(1-F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i}Y_j(t-1)}{d_i}\right),\ 1-F_\epsilon\left(-v_{i'}-\beta\frac{\sum_{j'\in\mathcal{N}_{i'}}Y_{j'}(t-1)}{d_{i'}}\right)\right)\\
\le\ &\frac{(L\beta)^2}{d_i d_{i'}}\mathrm{Cov}\left(\sum_{j\in\mathcal{N}_i}Y_j(t-1),\ \sum_{j'\in\mathcal{N}_{i'}}Y_{j'}(t-1)\right) \qquad (29a)\\
=\ &\frac{\rho^2}{d_i d_{i'}}\sum_{j\in\mathcal{N}_i}\sum_{j'\in\mathcal{N}_{i'}}\mathrm{Cov}\left(Y_j(t-1),\ Y_{j'}(t-1)\right). \qquad (29b)
\end{aligned}$$

where (29a) follows from Lemma 8 and (29b) follows from the definition of covariance of linear combinations and the fact that $\rho = L\beta$.

So far, we have proved that the covariance of $Y_i(t)$ and $Y_{i'}(t)$ for any pair (i,i') where $i,i'\in V$ and $i\neq i'$ can be upper bounded as follows:

$$\mathrm{Cov}(Y_i(t),Y_{i'}(t)) \le \frac{\rho^2}{d_i d_{i'}}\sum_{j\in\mathcal{N}_i}\sum_{j'\in\mathcal{N}_{i'}}\mathrm{Cov}\left(Y_j(t-1),\ Y_{j'}(t-1)\right). \qquad (30)$$

Meanwhile, for any $i\in V$, it is trivial that $\mathrm{Var}(Y_i(t))$ can be upper bounded by constant $1/4$ as $Y_i(t)$ is a binary variable.

When $t=1$, the covariance matrix can be upper bounded by $1/4\cdot\mathbf{I}$ as all the agents' decisions are not correlated. Recalling the definition of $\tilde{\mathbf{A}}$, iteratively, we can upper bound the covariance matrix $\boldsymbol{\Sigma}(t)$ as

$$\boldsymbol{\Sigma}(t) \le \frac{1}{4}\mathbf{I}+\rho^2\tilde{\mathbf{A}}\boldsymbol{\Sigma}(t-1)\tilde{A}^\top \le \frac{1}{4}\mathbf{I}+\rho^2\tilde{\mathbf{A}}\left[\frac{1}{4}\mathbf{I}+\rho^2\tilde{\mathbf{A}}\boldsymbol{\Sigma}(t-2)\tilde{\mathbf{A}}^\top\right]\tilde{\mathbf{A}}^\top \le \ldots \le \frac{1}{4}\left[\mathbf{I}+\sum_{\tau=1}^{t-1}\rho^{2\tau}\tilde{\mathbf{A}}^\tau\left(\tilde{\mathbf{A}}^\top\right)^\tau\right],$$

where the inequalities follow since (30) and the trivial $1/4$ bound for the variance of binary variables. □

Proof of Lemma 2: Based on Lemma 1, we can upper bound $\kappa_i(t)$ by the i-th diagonal entry of matrix

$$\tilde{\mathbf{A}}\boldsymbol{\Sigma}(t)\tilde{\mathbf{A}}^\top = \frac{1}{4}\left[\sum_{\tau=1}^{t}\rho^{2\tau-2}\tilde{\mathbf{A}}^\tau\left(\tilde{\mathbf{A}}^\top\right)^\tau\right].$$

Hereafter, we first reformulate a closed-form expression for this upper bound and provide a relaxation for this upper bound that can be calculated much easier.

Let $\mathbf{e}_i$ denote the vector when only the i-th entry equals 1 and other entries are 0. For an arbitrary matrix $\mathbf{M}$, the i-th diagonal entry of matrix $\mathbf{M}\mathbf{M}^\top$ equals to

$$\mathbf{e}_i^\top\mathbf{M}\mathbf{M}^\top\mathbf{e}_i = \sum_{j=1}^{n}M_{ij}^2.$$

Therefore, we can represent the vector of all diagonal entries to be $(\mathbf{M}\odot\mathbf{M})\,\mathbf{e}$, where $\odot$ means Hadamard product. As a consequence, we arrive at

$$\boldsymbol{\kappa}(t) \le \frac{1}{4}\left[\sum_{\tau=1}^{t}\rho^{2\tau-2}\tilde{\mathbf{A}}^\tau\odot\tilde{\mathbf{A}}^\tau\right]\mathbf{e}.$$

This upper bound involves the sum of a series of vectors, which cannot be reduced to a simplified expression. For a better interpretation of our results, we further relate it to Neumann series, so that it can be represented by a much compact form. Specifically, we will show that, for all $t \geq 0$,

$$\tilde{\mathbf{A}} \left(\tilde{\mathbf{A}}^\top\right)^t \leq \mathbf{b}\mathbf{e}^\top.$$

We will show that this inequality holds by induction.

Base case: $t = 1$. For all $i, j \in V$,

$$\left(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top\right)_{ij} = \frac{1}{d_i d_j} \sum_{k=1}^{n} \mathbb{1}\{k \text{ is the common in-neighbors of } i \text{ and } j\} \leq \frac{1}{d_i},$$

where the inequality follows since the number of common in-neighbors of agents i and j is smaller than d_j. As a result, $\tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top \leq \mathbf{b}\mathbf{e}^\top$.

To show time $t = s+1$: Assume that $\tilde{\mathbf{A}} \left(\tilde{\mathbf{A}}^\top\right)^s \leq \mathbf{b}\mathbf{e}^\top$. We have

$$\mathbf{b}\mathbf{e}^\top \tilde{\mathbf{A}}^\top \leq \mathbf{b}\mathbf{e}^\top,$$

where the inequality follows since $\tilde{\mathbf{A}}$ is a row-stochastic matrix. Hence, we can then prove that

$$\tilde{\mathbf{A}} \left(\tilde{\mathbf{A}}^\top\right)^s \leq \mathbf{b}\mathbf{e}^\top \left(\tilde{\mathbf{A}}^\top\right)^{s-1} \leq \mathbf{b}\mathbf{e}^\top,$$

where the inequality follows since all entries of $\tilde{\mathbf{A}}$ and $\mathbf{b}\mathbf{e}^\top$ are nonnegative.

Consequently, we have

$$\mathbf{e}_i^\top \tilde{\mathbf{A}}^\tau \left(\tilde{\mathbf{A}}^\top\right)^\tau \mathbf{e}_i \leq \mathbf{e}_i^\top \tilde{\mathbf{A}}^{\tau-1} \mathbf{b}\mathbf{e}^\top \mathbf{e}_i = \mathbf{e}_i^\top \tilde{\mathbf{A}}^{\tau-1} \mathbf{b}, \tag{31}$$

where the inequality follows since all entries of $\tilde{\mathbf{A}}$ and $\mathbf{e}_i$ are nonnegative.

In conclusion, we can derive the relaxed upper bound for $\boldsymbol{\kappa}(t)$ as

$$\boldsymbol{\kappa}(t) \leq \frac{1}{4}\left[\sum_{\tau=1}^{t} \rho^{2\tau-2} \tilde{\mathbf{A}}^\tau \odot \left(\tilde{\mathbf{A}}^\top\right)^\tau\right] \mathbf{e} \leq \frac{1}{4}\left[\sum_{\tau=1}^{t} \rho^{2\tau-2} \tilde{\mathbf{A}}^{\tau-1}\right] \mathbf{b} = \frac{1}{4}\left[\mathbf{I} + \sum_{\tau=1}^{t-1} \rho^{2\tau} \tilde{\mathbf{A}}^\tau\right] \mathbf{b},$$

where the second inequality follows from (31). □

B.3. Supporting Arguments for Theorem 1

Proof of Lemma 3: Let $\Delta_i(t) = \frac{\beta}{d_i}\left(\sum_{j\in\mathcal{N}_i} q_j(t-1) - \sum_{j\in\mathcal{N}_i} Y_j(t-1)\right)$. For any $i \in V$ and $t \geq 0$, the adoption probability of agent i at time t can be written as

$$\begin{aligned} q_i(t) &= \mathbb{E}_{\mathbf{Y}(t-1)}\left[\mathbb{E}_{\boldsymbol{\epsilon}(t)}\left[Y_i(t) \,\middle|\, \mathbf{Y}(t-1)\right]\right] = \mathbb{E}_{\mathbf{Y}(t-1)}\left[1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i}\right)\right] \\ &= 1 - \mathbb{E}_{\mathbf{Y}(t-1)}\left[F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1)\right)\right]. \end{aligned}$$

Therefore, we have

$$\begin{aligned} &\left|\mathbb{E}_{\mathbf{Y}(t-1)}\left[F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1)\right) - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i}\right)\right]\right| \\ = &\sqrt{\left(\mathbb{E}_{\mathbf{Y}(t-1)}\left[F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1)\right) - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i}\right)\right]\right)^2} \end{aligned}$$

$$\leq \sqrt{\mathbb{E}_{\mathbf{Y}(t-1)}\left[\left(F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i}q_j(t-1)}{d_i}+\Delta_i(t-1)\right)-F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i}q_j(t-1)}{d_i}\right)\right)^2\right]}$$

$$\leq \sqrt{\mathbb{E}_{\mathbf{Y}(t-1)}\left[L^2\left|\Delta_i(t-1)\right|^2\right]} = \sqrt{\mathbb{E}_{\mathbf{Y}(t-1)}\left[(L\beta)^2\left(\frac{\sum_{j\in\mathcal{N}_i}q_j(t-1)}{d_i}-\frac{\sum_{j\in\mathcal{N}_i}Y_j(t-1)}{d_i}\right)^2\right]}$$

$$= \sqrt{\rho^2\,\mathrm{Var}\left[\frac{\sum_{j\in\mathcal{N}_i}Y_j(t-1)}{d_i}\right]},$$

where the first inequality follows by Jensen's inequality and the second inequality follows by Assumption 1.

Since $\boldsymbol{\delta}=\frac{\rho}{2}\left[\mathcal{B}(G;\rho)\right]^{\frac{1}{2}}$, by Lemma 2, we can obtain

$$\left|\mathbb{E}_{\mathbf{Y}(t-1)}\left[F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i}q_j(t-1)}{d_i}+\Delta_i(t-1)\right)-F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i}q_j(t-1)}{d_i}\right)\right]\right| \leq \delta_i,$$

which further leads to

$$1-F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i}q_j(t-1)}{d_i}\right)-\delta_i \leq q_i(t) \leq 1-F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i}q_j(t-1)}{d_i}\right)+\delta_i.$$

In summary, we have

$$\left|\mathbf{h}\big(\mathbf{q}(t-1)\big)-\mathbf{q}(t)\right|_{\mathrm{ew}} \leq \boldsymbol{\delta} = \frac{\rho}{2}\left[\mathcal{B}(G;\rho)\right]^{\frac{1}{2}}_{\mathrm{ew}} \leq \frac{\rho}{2}\left[\left(\mathbf{I}-\rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}\right]^{\frac{1}{2}}_{\mathrm{ew}}.$$

and this concludes the proof. □

Finally, we prove our main result, i.e., Theorem 1.

Proof of Theorem 1: We first show by induction that, $\underline{\boldsymbol{\mu}}(t)\leq\mathbf{q}(t)\leq\overline{\boldsymbol{\mu}}(t)$ for each $t\geq 0$.

Base case: $\boldsymbol{t=0}$. By definition, we have $\underline{\boldsymbol{\mu}}(0)=\mathbf{q}(0)=\overline{\boldsymbol{\mu}}(0)$.

To show $\boldsymbol{t=s+1}$: Assume that $\underline{\boldsymbol{\mu}}(t)\leq\mathbf{q}(t)\leq\overline{\boldsymbol{\mu}}(t)$. Then we have

$$\underline{\boldsymbol{\mu}}(s+1) = \mathbf{h}_{-\boldsymbol{\delta}}\big(\underline{\boldsymbol{\mu}}(s)\big) \leq \mathbf{h}_{-\boldsymbol{\delta}}\big(\mathbf{q}(s)\big) \leq \mathbf{q}(s+1) \leq \mathbf{h}_{\boldsymbol{\delta}}\big(\mathbf{q}(s)\big) \leq \mathbf{h}_{\boldsymbol{\delta}}\big(\overline{\boldsymbol{\mu}}(s)\big) = \overline{\boldsymbol{\mu}}(s+1),$$

where the first and last inequalities follow from Proposition 2(i) while the other two follow from Lemma 3.

Also, clearly Proposition 2 implies that $\underline{\boldsymbol{\mu}}(t)\leq\boldsymbol{\mu}(t)\leq\overline{\boldsymbol{\mu}}(t)$ for all $t\geq 0$. By the contraction mapping theorem, we know that $\underline{\boldsymbol{\mu}}(t)$ (resp. $\overline{\boldsymbol{\mu}}(t)$) converges to $\underline{\boldsymbol{\mu}}^*$ (resp. $\overline{\boldsymbol{\mu}}^*$) where $\underline{\boldsymbol{\mu}}^*$ (resp. $\overline{\boldsymbol{\mu}}^*$) is the fixed-point solution for $\mathbf{h}_{-\boldsymbol{\delta}}(\underline{\boldsymbol{\mu}}^*)=\underline{\boldsymbol{\mu}}^*$ (resp. $\mathbf{h}_{\boldsymbol{\delta}}(\overline{\boldsymbol{\mu}}^*)=\overline{\boldsymbol{\mu}}^*$). Thus, the following result holds,

$$\underline{\boldsymbol{\mu}}^*\leq\mathbf{q}^*\leq\overline{\boldsymbol{\mu}}^* \text{ and } \underline{\boldsymbol{\mu}}^*\leq\boldsymbol{\mu}^*\leq\overline{\boldsymbol{\mu}}^*. \tag{32}$$

By the definition of AEOs, the difference between $\underline{\boldsymbol{\mu}}^*$ and $\overline{\boldsymbol{\mu}}^*$ can be written as

$$\overline{\boldsymbol{\mu}}^*-\underline{\boldsymbol{\mu}}^* = \mathbf{h}(\overline{\boldsymbol{\mu}}^*)-\mathbf{h}(\underline{\boldsymbol{\mu}}^*)+2\boldsymbol{\delta}.$$

Let $\Delta\boldsymbol{\mu}=\overline{\boldsymbol{\mu}}^*-\underline{\boldsymbol{\mu}}^*$. For all $i\in V$,

$$|\Delta\mu_i| \leq \rho\left|\frac{\sum_{j\in\mathcal{N}_i}\overline{\mu}^*_j}{d_i}-\frac{\sum_{j\in\mathcal{N}_i}\underline{\mu}^*_j}{d_i}\right|+2\delta_i = \rho\left|\frac{\sum_{j\in\mathcal{N}_i}\Delta\mu_i}{d_i}\right|+2\delta_i, \tag{33}$$

where the inequality comes from Assumption 1.

In matrix form, we can write it as $|\Delta\boldsymbol{\mu}|_{\mathrm{ew}}\leq\rho\tilde{\mathbf{A}}|\Delta\boldsymbol{\mu}|_{\mathrm{ew}}+2\boldsymbol{\delta}$ or equivalently

$$(\mathbf{I}-\rho\tilde{\mathbf{A}})|\Delta\boldsymbol{\mu}|_{\mathrm{ew}}\leq 2\boldsymbol{\delta}. \tag{34}$$

Recall that the inverse matrix $(\mathbf{I}-\rho\tilde{\mathbf{A}})^{-1}$ can be expanded into the Neumann series $\mathbf{I}+\sum_{\ell=1}^{\infty}\rho^{\ell}\tilde{\mathbf{A}}^{\ell}$. Given that all elements of $\tilde{\mathbf{A}}$ are non-negative, it follows that all elements of $(\mathbf{I}-\rho\tilde{\mathbf{A}})^{-1}$ are also non-negative. Therefore, when we pre-multiply both sides of (34) by $(\mathbf{I}-\rho\tilde{\mathbf{A}})^{-1}$, we obtain the inequality

$$|\Delta\boldsymbol{\mu}|_{\text{ew}} \;\le\; 2\left(\mathbf{I}-\rho\tilde{\mathbf{A}}\right)^{-1}\boldsymbol{\delta}. \tag{35}$$

Combining (32) and (35), we finally have the following chain of inequalities:

$$\begin{aligned}
|\mathbf{q}^*-\boldsymbol{\mu}^*|_{\text{ew}} \;&\le\; |\Delta\boldsymbol{\mu}|_{\text{ew}} \;\le\; \rho(\mathbf{I}-\rho\tilde{\mathbf{A}})^{-1}\left[\mathcal{B}(G;\rho)\right]_{\text{ew}}^{\frac{1}{2}} \;\le\; \rho\left(\mathbf{I}-\rho\tilde{\mathbf{A}}\right)^{-1}\left[\left(\mathbf{I}-\rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}\right]_{\text{ew}}^{\frac{1}{2}}\\
&=\; \frac{\rho}{1-\rho}\left[(1-\rho)\left(\mathbf{I}-\rho\tilde{\mathbf{A}}\right)^{-1}\right]\cdot\left[\left(\mathbf{I}-\rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}\right]_{\text{ew}}^{\frac{1}{2}} \;\le\; \frac{\rho}{1-\rho}\left[(1-\rho)\left(\mathbf{I}-\rho\tilde{\mathbf{A}}\right)^{-1}\left(\mathbf{I}-\rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}\right]_{\text{ew}}^{\frac{1}{2}}\\
&=\; \frac{\rho}{\sqrt{1-\rho}}\left[\left(\sum_{s=0}^{\infty}\rho^s\tilde{\mathbf{A}}^{\,s}\right)\left(\sum_{t=0}^{\infty}\rho^{2t}\tilde{\mathbf{A}}^{\,t}\right)\mathbf{b}\right]_{\text{ew}}^{\frac{1}{2}} \;=\; \frac{\rho}{\sqrt{1-\rho}}\left[\sum_{\ell=0}^{\infty}\left(\sum_{s,t\in\mathbb{Z}_+:\;s+t=\ell}\rho^{s+2t}\right)\tilde{\mathbf{A}}^{\,\ell}\mathbf{b}\right]_{\text{ew}}^{\frac{1}{2}}\\
&=\; \frac{\rho}{\sqrt{1-\rho}}\left[\sum_{\ell=0}^{\infty}\frac{\rho^{\ell+1}-\rho^{2\ell+2}}{\rho-\rho^2}\tilde{\mathbf{A}}^{\,\ell}\mathbf{b}\right]_{\text{ew}}^{\frac{1}{2}} \;\le\; \frac{\rho}{\sqrt{1-\rho}}\left[\sum_{\ell=0}^{\infty}\frac{\rho^{\ell}}{1-\rho}\tilde{\mathbf{A}}^{\,\ell}\mathbf{b}\right]_{\text{ew}}^{\frac{1}{2}} \;=\; \frac{\rho}{1-\rho}\left[\left(\mathbf{I}+\sum_{\ell=1}^{\infty}\rho^{\ell}\tilde{\mathbf{A}}^{\,\ell}\right)\mathbf{b}\right]_{\text{ew}}^{\frac{1}{2}}.
\end{aligned}$$

where the fourth inequality follows from Jensen's inequality provided $(1-\rho)\left(\mathbf{I}-\rho\tilde{\mathbf{A}}\right)^{-1}$ is a row-stochastic matrix. This concludes the proof.

We also remark that the bounds above also apply for each $t\ge 0$. Indeed, $\underline{\boldsymbol{\mu}}(t)\le\mathbf{q}(t),\boldsymbol{\mu}(t)\le\overline{\boldsymbol{\mu}}(t)$. Thus, it suffices to bound $\Delta\boldsymbol{\mu}(t)=\overline{\boldsymbol{\mu}}(t)-\underline{\boldsymbol{\mu}}(t)$. An argument similar to (33) shows that $|\Delta\boldsymbol{\mu}(t)|_{\text{ew}}\le\rho\tilde{\mathbf{A}}|\Delta\boldsymbol{\mu}(t-1)|_{\text{ew}}+2\boldsymbol{\delta}$. Clearly, $|\Delta\boldsymbol{\mu}(0)|_{\text{ew}}=0\le 2\left(\mathbf{I}-\rho\tilde{\mathbf{A}}\right)^{-1}\boldsymbol{\delta}$. Assume that $|\Delta\boldsymbol{\mu}(t-1)|_{\text{ew}}\le 2\left(\mathbf{I}-\rho\tilde{\mathbf{A}}\right)^{-1}\boldsymbol{\delta}$. Then,

$$\begin{aligned}
|\Delta\boldsymbol{\mu}(t)|_{\text{ew}} \;&\le\; \rho\tilde{\mathbf{A}}|\Delta\boldsymbol{\mu}(t-1)|_{\text{ew}}+2\boldsymbol{\delta} \;\le\; \rho\tilde{\mathbf{A}}\cdot 2\left(\mathbf{I}-\rho\tilde{\mathbf{A}}\right)^{-1}\boldsymbol{\delta}+2\boldsymbol{\delta}\\
&=\; -\left(\mathbf{I}-\rho\tilde{\mathbf{A}}\right)\cdot 2\left(\mathbf{I}-\rho\tilde{\mathbf{A}}\right)^{-1}\boldsymbol{\delta}+2\boldsymbol{\delta}+2\left(\mathbf{I}-\rho\tilde{\mathbf{A}}\right)^{-1}\boldsymbol{\delta} \;=\; 2\left(\mathbf{I}-\rho\tilde{\mathbf{A}}\right)^{-1}\boldsymbol{\delta},
\end{aligned}$$

as desired. We conclude the proof. □

In the following, we prove the corollary for Theorem 1.

Proof of Corollary 2: From Theorem 1, it holds that

$$\begin{aligned}
\frac{1}{n}\|\mathbf{q}^*-\boldsymbol{\mu}^*\|_1 \;&\le\; \frac{C_\rho}{n}\mathbf{e}^\top\left[\mathcal{C}(G;\rho)\right]_{\text{ew}}^{\frac{1}{2}}\\
&\overset{(a)}{\le}\; \frac{C_\rho}{\sqrt{n}}\sqrt{\left\|\left[\mathcal{C}(G;\rho)\right]_{\text{ew}}^{\frac{1}{2}}\right\|_1} \;=\; \frac{C_\rho\sqrt{1-\rho}}{\sqrt{n}}\sqrt{\mathbf{e}^\top\left(\mathbf{I}+\sum_{\ell=1}^{\infty}\rho^{\ell}\tilde{\mathbf{A}}^{\,\ell}\right)\mathbf{b}}\,,
\end{aligned}$$

where (a) follows due to Cauchy-Schwarz inequality. Further, (a) proves (9) in the corollary.

In the following, we will bound $\mathbf{e}^\top\left(\mathbf{I}+\sum_{\ell=1}^{\infty}\rho^{\ell}\tilde{\mathbf{A}}^{\,\ell}\right)\mathbf{b}$. Let us define $\mathbf{D}=\text{diag}(\mathbf{b})$, so that $\tilde{\mathbf{A}}=\mathbf{D}\mathbf{A}^\top$ holds, where $\mathbf{A}$ is the adjacency matrix. Further, we define

$$\mathbf{Q}(s) \;:=\; \mathbf{A}^\top\tilde{\mathbf{A}}^{\,s-1} \;=\; \mathbf{A}^\top\left(\mathbf{D}\mathbf{A}^\top\right)^{s-1}, \quad \forall\; s\ge 0.$$

Then, it holds that

$$\begin{aligned}
&\left\|\tilde{\mathbf{A}}^{\,s}\mathbf{b}\right\|_1 = \mathbf{e}^\top\mathbf{D}\mathbf{Q}(s)\mathbf{D}\mathbf{e}\\
&=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{d_i d_j}Q_{ij}(s)\le\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2}\left(\frac{1}{d_i^2}+\frac{1}{d_j^2}\right)Q_{ij}(s)=\frac{1}{2}\left\|\mathbf{D}^2\mathbf{Q}(s)\mathbf{e}\right\|_1+\frac{1}{2}\left\|\mathbf{Q}(s)\mathbf{D}^2\mathbf{e}\right\|_1,
\end{aligned} \tag{36}$$

where the inequality follows from the AM-GM inequality.

We then bound the two terms in (36) as follows:

$$\frac{1}{2}\left\|\mathbf{D}^2\mathbf{Q}(s)\mathbf{e}\right\|_1 = \frac{1}{2}\left\|\mathbf{D}^2\mathbf{A}^\top\tilde{\mathbf{A}}^{s-1}\mathbf{e}\right\|_1 = \frac{1}{2}\left\|\mathbf{D}\mathbf{D}\mathbf{A}^\top\mathbf{e}\right\|_1 = \frac{1}{2}\left\|\mathbf{D}\mathbf{e}\right\|_1 = \frac{1}{2}\left\|\mathbf{b}\right\|_1,$$

where the second and the third equalities follow because $\tilde{\mathbf{A}}$ is row-stochastic, and

$$\frac{1}{2}\left\|\mathbf{Q}(s)\mathbf{D}^2\mathbf{e}\right\|_1 = \frac{1}{2}\left\|\left(\mathbf{A}^\top\mathbf{D}\right)^s\mathbf{D}\mathbf{e}\right\|_1 \le \frac{1}{2}\left\|\mathbf{A}^\top\mathbf{D}\right\|_1^s\left\|\mathbf{D}\mathbf{e}\right\|_1 \overset{(b)}{=} \frac{1}{2}r^s(G)\left\|\mathbf{D}\mathbf{e}\right\|_1 = \frac{1}{2}r^s(G)\left\|\mathbf{b}\right\|_1,$$

where the inequality follows from the definition of matrix-ℓ_1-norm as an operator norm and (b) follows since by inspection $r(G)=\mathbf{A}^\top\mathbf{D}$. Combined with (36), we get

$$\left\|\tilde{\mathbf{A}}^s\mathbf{b}\right\|_1 \le \frac{1}{2}\left(1+r^s(G)\right)\left\|\mathbf{b}\right\|_1 \le r^s(G)\left\|\mathbf{b}\right\|_1, \tag{37}$$

where the last inequality follows since by definition the largest out-in-degree ratio $r(G)\ge 1$. Therefore, as long as $\rho r(G)<1$,

$$\mathbf{e}^\top\left(\mathbf{I}+\sum_{\ell=1}^{\infty}\rho^\ell\tilde{\mathbf{A}}^\ell\right)\mathbf{b} = \left\|\left(\mathbf{I}+\sum_{\ell=1}^{\infty}\rho^\ell\tilde{\mathbf{A}}^\ell\right)\mathbf{b}\right\|_1 \le \left\|\mathbf{b}\right\|_1+\sum_{\ell=1}^{\infty}\rho^\ell\left\|\tilde{\mathbf{A}}^\ell\mathbf{b}\right\|_1 \le \frac{1}{1-\rho r(G)}\left\|\mathbf{b}\right\|_1,$$

where the first inequality follows from the subadditivity of norms and the last inequality follows from (37). This concludes the proof. □

Appendix C: Extensions to Non-i.i.d. Noise Assumption

In our basic model, we assume that the random noise $\epsilon_i(t)$ is i.i.d. for all $i\in V$ and $t\ge 1$. In this section, we explore extensions where this assumption is relaxed. The performance of the FPA scheme in the absence of the i.i.d. assumption depends on the specific definition of the joint distribution. We outline and discuss two cases: The first case considers the situation where the noise is not identically distributed across time, and the second case examines when the noise is not independently distributed across agents at the same time.

C.1. Non-identical Noise Assumption

We assume the random noise $\epsilon_i(t)$ for all $i\in V$ and $t\ge 1$ is independently drawn from distributions with varying parameters, but converges to a specific distribution as $t\to\infty$. Let $F_{\epsilon(t)}(\cdot)$ denote the CDF of the random noise at time t, and let $F_\epsilon(\cdot)$ denote the CDF of the asymptotic distribution as $t\to\infty$. For instance, if we assume $\epsilon_i(t)$ follows a Logistic distribution, we can model it as $\epsilon_i(t)\sim\text{logistic}(0,\sigma^2(t))$ for all agent $i\in V$ at time step $t\ge 1$, where $\sigma(t)$ diminishes over time and eventually converges to a constant σ as $t\to\infty$.

In this setting, the diffusion process can be viewed as a nonstationary Markov chain, where the transition matrix evolves over time. However, since the transition matrix converges to the one constructed using the asymptotic distribution, the system will eventually behave as though the noise distribution is fixed at the asymptotic distribution, leading to a stable long-term behavior.

As a result, the FPA scheme can be easily generalized from the i.i.d. case by changing the parameter L to denote the maximum Lipschitz constant of all the CDFs $F_{\epsilon(t)}(\cdot)$ for $t\ge 1$. Specifically, as long as Assumption 1 is satisfied, we can still construct the FPA solution as in (5), and the approximation error can be bounded as shown in Theorem 1.

Note that requiring L to be the maximum Lipschitz constant is not a stringent condition. For example, in the case of the Logistic distribution with diminishing variance, L just corresponds to the Lipschitz constant of asymptotic distribution logistic$(0, \sigma^2)$. Therefore, the FPA scheme and the general approximation error analysis remain the same as in the original i.i.d. case. We also want to make the following remark: As explained in Section 2.1 following Assumption 1, the bound we provide is instance-independent, meaning it assumes the intrinsic value $\mathbf{v}$ can take arbitrary values. If we consider an instance-dependent bound for a specific $\mathbf{v}$, the parameter L should be taken as the maximum Lipschitz constant only within the domain $[-\max_{i\in V} v_i - \beta, -\min_{i\in V} v_i]$ of all CDFs. In this case, L may not necessarily correspond to the Lipschitz constant of the asymptotic distribution. As a result, introducing nonstationarity (e.g., time-varying noise) could increase the error bound compared to the stationary case.

C.2. Non-independent Noise Assumption

When the random noise $\epsilon_i(t)$ is correlated across the network, the problem becomes significantly more complex. The primary challenge is to derive a counterpart to Lemma 2 that accounts for the correlation in the random noise and provides an upper bound for the variance of the local network effect.

We start by defining the correlation across the random noise. Due to the nonlinear transitions and correlated adoption states, it is insufficient to specify just the covariance matrix of the random noise. We must also capture a more specific structure of this correlation. To do so, we introduce a quadrant function $H(x, y)$, which quantifies the joint probability of $\epsilon_i(t)$ and $\epsilon_j(t)$ as follows:

$$H(x, y) := \mathbb{P}(\epsilon_i(t) \le x, \epsilon_j(t) \le y) - \mathbb{P}(\epsilon_i(t) \le x)\mathbb{P}(\epsilon_j(t) \le y).$$

To quantify the magnitude of the correlation, we introduce a parameter a that represents the uniform upper bound of $H(x, y)$, i.e., for all $x, y \in \mathbb{R}$, we require $0 \le H(x, y) \le a$. Notably, a is naturally bounded by $1/4$, since:

$$\begin{aligned} H(x,y) &= \mathbb{P}(\epsilon_i(t) \le x, \epsilon_j(t) \le y) - \mathbb{P}(\epsilon_i(t) \le x)\mathbb{P}(\epsilon_j(t) \le y) = \mathrm{Cov}\left(\mathbb{1}\{\epsilon_i(t) \le x\}, \mathbb{1}\{\epsilon_j(t) \le y\}\right) \\ &\le \sqrt{\mathrm{Var}\left(\mathbb{1}\{\epsilon_i(t) \le x\}\right)\mathrm{Var}\left(\mathbb{1}\{\epsilon_j(t) \le y\}\right)} \\ &\le \frac{1}{4}, \end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality holds because the variance of a binary random variable is upper bounded by $1/4$.

Next, we present Lemma 9, which provides an upper bound for the variance of the local network effect when the random noise may be correlated at a single time step.

LEMMA 9 **(Local Network Effect Variance)**. *Under Assumptions 1, 2, and $0 \le H(\cdot, \cdot) \le a \le 1/4$, for any diffusion instance $\left(G, \mathbf{v}, F_\epsilon(\cdot), \beta\right)$ and $t \ge 1$, the variance of network effect can be upper bounded by*

$$\boldsymbol{\kappa}(t) \le \left(\frac{1}{4} - a\right)\mathcal{B}(G; \rho) + \frac{a}{1-\rho^2}\mathbf{e} \le \left(\frac{1}{4} - a\right)\left(\mathbf{I} - \rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b} + \frac{a}{1-\rho^2}\mathbf{e}.$$

We observe that when the random noise is correlated, the upper bound for the variance of the local network effect increases, becoming a linear combination of the original bound (under the independent noise assumption) and a vector of ones. When $a = 0$, corresponding to the independent noise case, the result in Lemma 9 reduces to the original result in Lemma 2. Using this result, we can derive the associated approximation error for the FPA scheme. Since the procedure follows similarly to the independent case, we omit a detailed demonstration here.

Proof of Lemma 9: While much of the proof follows the same structure as in Lemmas 1 and 2, the key difference arises in the decomposition of the covariance in (28).

When the random noise is correlated, the first term in (28) will no longer be zero, but it can be upper-bounded as follows:

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{Y}(t-1)}\left[\mathrm{Cov}_{\boldsymbol{\epsilon}(t)}\left(\mathbb{E}\left[Y_i(t)|\mathbf{Y}(t-1),\boldsymbol{\epsilon}(t)\right],\mathbb{E}\left[Y_{i'}(t)|\mathbf{Y}(t-1),\boldsymbol{\epsilon}(t)\right]\right)\middle|\mathbf{Y}(t-1)\right]\\
&=\mathbb{E}_{\mathbf{Y}(t-1)}\left[\mathrm{Cov}_{\boldsymbol{\epsilon}(t)}\left(\mathbb{1}\left\{\epsilon_i(t)\le -v_i-\beta\frac{\sum_{j\in\mathcal{N}_i}Y_j(t-1)}{d_i}\right\},\mathbb{1}\left\{\epsilon_{i'}(t)\le -v_{i'}-\beta\frac{\sum_{j'\in\mathcal{N}_{i'}}Y_{j'}(t-1)}{d_{i'}}\right\}\right)\middle|\mathbf{Y}(t-1)\right]\\
&=\mathbb{E}_{\mathbf{Y}(t-1)}\left[H\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i}Y_j(t-1)}{d_i},-v_{i'}-\beta\frac{\sum_{j'\in\mathcal{N}_{i'}}Y_{j'}(t-1)}{d_{i'}}\right)\middle|\mathbf{Y}(t-1)\right]\\
&\le a.
\end{aligned}
$$

Thus, the covariance matrix of the adoption states is upper bounded by:

$$
\begin{aligned}
\boldsymbol{\Sigma}(t)&\le\rho^2\tilde{\mathbf{A}}\boldsymbol{\Sigma}(t-1)\tilde{\mathbf{A}}^\top+\left(\frac{1}{4}-a\right)\mathbf{I}+a\mathbf{e}\mathbf{e}^\top\\
&\le\rho^4\tilde{\mathbf{A}}^2\boldsymbol{\Sigma}(t-2)(\tilde{\mathbf{A}}^\top)^2+\left(\frac{1}{4}-a\right)\left(\mathbf{I}+\rho^2\tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top\right)+a(1+\rho^2)\mathbf{e}\mathbf{e}^\top\le\cdots\\
&\le\left(\frac{1}{4}-a\right)\left(\mathbf{I}+\sum_{\tau=1}^{t-1}\rho^{2\tau}\tilde{\mathbf{A}}^\tau(\tilde{\mathbf{A}}^\top)^\tau\right)+a\left(1+\sum_{\tau=1}^{t-1}\rho^{2\tau}\right)\mathbf{e}\mathbf{e}^\top.
\end{aligned}
$$

Consequently, the variance of the local network effect can be upper bounded by

$$
\begin{aligned}
\boldsymbol{\kappa}(t)&\le\left(\frac{1}{4}-a\right)\left[\sum_{t=1}^{t}\rho^{2\tau-2}\tilde{\mathbf{A}}^\tau\odot(\tilde{\mathbf{A}}^\top)^\tau\right]+a\sum_{\tau=0}^{t-1}\rho^{2\tau}\mathbf{e}\\
&\le\left(\frac{1}{4}-a\right)\left[\mathbf{I}+\sum_{\tau=1}^{t-1}\rho^{2\tau}\tilde{\mathbf{A}}^\tau\right]\mathbf{b}+a\sum_{\tau=0}^{t-1}\rho^{2\tau}\mathbf{e}.
\end{aligned}
$$

In conclusion, we have

$$
\boldsymbol{\kappa}(t)\le\left(\frac{1}{4}-a\right)\mathcal{B}(G;\rho)+\frac{a}{1-\rho^2}\mathbf{e}\le\left(\frac{1}{4}-a\right)\left(\mathbf{I}-\rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}+\frac{a}{1-\rho^2}\mathbf{e}.
$$

□

Appendix D: Supporting Arguments for Section 4

For a refined upper bound on the approximation error, we first show the proof for Lemma 4, which is a refined version of Lemma 3.

Proof of Lemma 4: Let $X_i = -v_i - \beta \frac{1}{d_i} \sum_{j\in\mathcal{N}_i} Y_j(t-1)$ and $\nu_i = -v_i - \beta \frac{1}{d_i} \sum_{j\in\mathcal{N}_i} q_j(t-1)$ for all $i \in V$. For any $i \in V$ and $t \geq 1$, the adoption probability of agent i at t can be written as

$$q_i(t) \;=\; \mathbb{E}_{\mathbf{Y}(t-1)}\Big[\mathbb{E}\big[y_i(t) \mid \mathbf{Y}(t-1)\big]\Big] \;=\; \mathbb{E}_{\mathbf{Y}(t-1)}\left[1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i}\right)\right] \;=\; 1 - \mathbb{E}_{X_i}\big[F_\epsilon\left(X_i\right)\big].$$

With Assumption 3, we can apply Taylor expansion to $F_\epsilon\left(X_i\right)$ and get

$$\begin{aligned}\Big|\mathbb{E}_{X_i}\big[F_\epsilon\left(X_i\right) - F_\epsilon\left(\nu_i\right)\big]\Big| \;&=\; \left|\mathbb{E}_{X_i}\left[F_\epsilon\left(\nu_i\right) + f_\epsilon(\nu_i)(X_i-\nu_i) + \frac{1}{2}f'_\epsilon(\tilde{X}_i)(X_i-\nu_i)^2 - F_\epsilon\left(\nu_i\right)\right]\right| \\ &=\; \frac{1}{2}\left|\mathbb{E}_{X_i}\left[f'_\epsilon(\tilde{X}_i)(X_i-\nu_i)^2\right]\right|,\end{aligned} \tag{38}$$

where $\tilde{X}_i$ is a random variable such that $\tilde{X}_i$ lies in between the random variable X_i and ν_i.

Consequently, we can upper bound (38) by

$$\begin{aligned}\left|\mathbb{E}_{X_i}\left[F_\epsilon\left(X_i\right) - F_\epsilon\left(\nu_i\right)\right]\right| \;&=\; \frac{1}{2}\left|\mathbb{E}_{X_i}\left[f'_\epsilon(\tilde{X}_i)(X_i-\nu_i)^2\right]\right| \;\leq\; \frac{1}{2}\mathbb{E}_{X_i}\left[|f'_\epsilon(\tilde{X}_i)|(X_i-\nu_i)^2\right] \;\leq\; \frac{L_f}{2}\operatorname{Var}\left(X_i\right) \\ &=\; \frac{L_f\beta^2}{2}\operatorname{Var}\left(\frac{1}{d_i}\sum_{j\in\mathcal{N}_i} Y_j(t-1)\right),\end{aligned}$$

where the first inequality comes from Jensen's inequality and the second inequality is from Assumption 3. Let $\boldsymbol{\eta} = \frac{L_f\beta^2}{8}\mathcal{B}(G;\rho)$. By applying Lemma 2, we can finally get

$$\Big|\mathbb{E}\big[F_\epsilon(X_i) - F_\epsilon\left(\nu_i\right)\big]\Big| \;\leq\; \eta_i,$$

which further leads to

$$1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i}\right) - \eta_i \;\leq\; q_i(t) \;\leq\; 1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i}\right) + \eta_i.$$

In conclusion, we have

$$\mathbf{h}(\mathbf{q}(t-1)) - \boldsymbol{\eta} \;\leq\; \mathbf{q}(t) \;\leq\; \mathbf{h}(\mathbf{q}(t-1)) + \boldsymbol{\eta}.$$

Moreover, because $\mathcal{B}(G;\rho) \leq \left(\mathbf{I} - \rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}$ as proved in Lemma 2, we can also naturally have

$$\Big|\mathbf{h}(\mathbf{q}(t-1)) - \boldsymbol{q}(t)\Big|_{\text{ew}} \leq \frac{L_f\beta^2}{8}\mathcal{B}(G;\rho) \leq \frac{L_f\beta^2}{8}\left(\mathbf{I} - \rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}.$$

□

Based on Lemma 4, we then show the proof for the refined Theorem 2 and Corollary 3.

Proof of Theorem 2: Following the same steps leading to (35), with Lemma 4, we obtain

$$|\Delta\boldsymbol{\mu}|_{\text{ew}} \;\leq\; 2\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1}\boldsymbol{\eta}. \tag{39}$$

Therefore, following the same line of analysis in the proof of Theorem 1, it holds that

$$\begin{aligned}|\mathbf{q}^* - \boldsymbol{\mu}^*|_{\text{ew}} \;&\leq\; |\Delta\boldsymbol{\mu}|_{\text{ew}} \;\leq\; 2\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1}\boldsymbol{\eta} \;=\; \frac{L_f\beta^2}{4}\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1}\cdot\mathcal{B}_(G;\rho) \\ &\leq\; \frac{L_f\beta^2}{4}\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1}\cdot\left(\mathbf{I} - \rho^2\tilde{\mathbf{A}}\right)^{-1}\mathbf{b} \;=\; \frac{L_f\beta^2}{4}\left[\left(\sum_{s=0}^{\infty}\rho^s\tilde{\mathbf{A}}^{\,s}\right)\left(\sum_{t=0}^{\infty}\rho^{2t}\tilde{\mathbf{A}}^{\,t}\right)\mathbf{b}\right] \\ &\leq\; \frac{L_f\beta^2}{4(1-\rho)}\cdot\left(\mathbf{I} + \sum_{\ell=1}^{\infty}\rho^\ell\tilde{\mathbf{A}}^{\,\ell}\right)\mathbf{b} \;=\; \frac{L_f\beta^2}{4(1-\rho)^2}\cdot\mathcal{C}(G;\rho).\end{aligned}$$

We conclude the proof. □

Proof of Corollary 3: By Theorem 2, we can upper bound the scaled ℓ_1-norm as

$$\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\le\; \frac{\tilde{C}}{n}\mathbf{e}^\top \mathcal{C}(G;\rho) \;=\; \frac{(1-\rho)\tilde{C}}{n}\cdot \mathbf{e}^\top \left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell\right)\mathbf{b}$$

Following the proof of Corollary 2, the last term can be bounded by

$$\frac{(1-\rho)\tilde{C}}{n}\cdot \mathbf{e}^\top \left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell\right)\mathbf{b} \;\le\; \frac{(1-\rho)\tilde{C}}{n(1-\rho r(G))}\,\|\mathbf{b}\|_1\,,$$

and we conclude the proof. □

LEMMA 10. *Given an entrywise increasing sequence of vectors* $\{\boldsymbol{\eta}(1), \boldsymbol{\eta}(2), \ldots\}$, *with* $\lim_{t\to\infty}\boldsymbol{\eta}(t) = \boldsymbol{\eta}$, *the dynamical system*

$$\boldsymbol{\mu}(t) = \mathbf{h}_{\boldsymbol{\eta}(t-1)}(\boldsymbol{\mu}(t-1)),\ \forall t \ge 1,$$

converges to $\boldsymbol{\nu}^*$ *when* t *tends to infinity, where* $\boldsymbol{\nu}^*$ *is the fixed-point solution of the system of equations* $\mathbf{h}_{\boldsymbol{\eta}}(\boldsymbol{\nu}) = \boldsymbol{\nu}$.

Proof of Lemma 10: We first define a set of auxiliary dynamical systems, including an upper bound dynamical system $\boldsymbol{\nu}(t)$, and for any $\tau \ge 0$, a set of lower bound dynamical system $\boldsymbol{\mu}^{[\tau]}(t)$ as follows:

$$\nu_i(t) = \begin{cases} q_i(0) & t = 0 \\ 1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i}\nu_j(t-1)}{d_i}\right) + \eta_i & t > 0 \end{cases}, \quad \text{for all } i \in V,$$

$$\mu_i^{[\tau]}(t) = \begin{cases} q_i(0) & t = 0 \\ 1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i}\mu_j^{[\tau]}(t-1)}{d_i}\right) + \eta_i(\min\{t-1,\tau\}) & t > 0 \end{cases}, \quad \text{for all } i \in V.$$

It is easy to verify that $\boldsymbol{\mu}^{[\tau]}(t) \le \boldsymbol{\mu}(t) \le \boldsymbol{\nu}(t)$. Then, by the contraction mapping theorem, we know that $\boldsymbol{\mu}^{[\tau]}(t)$ (resp. $\boldsymbol{\nu}(t)$) converges to $\boldsymbol{\mu}^{[\tau]*}$ (resp. $\boldsymbol{\nu}^*$) where $\boldsymbol{\mu}^{[\tau]*}$ (resp. $\boldsymbol{\nu}^*$) is the fixed-point solution for the system of equations $\mathbf{h}_{\boldsymbol{\eta}(\tau)}(\boldsymbol{\mu}^{[\tau]}) = \boldsymbol{\mu}^{[\tau]}$ (resp. $\mathbf{h}_{\boldsymbol{\eta}}(\boldsymbol{\nu}) = \boldsymbol{\nu}$). Thus, the following result holds,

$$\boldsymbol{\mu}^{[\tau]*} \le \boldsymbol{\mu}^* \le \boldsymbol{\nu}^*.$$

By the proof of Theorem 1, we know that the sequence $\{\boldsymbol{\mu}^{[0]*}, \boldsymbol{\mu}^{[1]*}, \cdots\}$ is also entrywise increasing. Hence, by monotone convergence theorem, we have

$$\lim_{\tau\to\infty}\boldsymbol{\mu}^{[\tau]*} = \boldsymbol{\nu}^*.$$

Finally, by squeeze theorem, we can conclude that $\boldsymbol{\mu}^* = \boldsymbol{\nu}^*$. □

Proof of Theorem 3: We use the diffusion instance given in the main text to show the lower bound. We first remark on the following facts that will be used in the next. For this specific instance, the CDF F_ϵ, PDF f_ϵ and the derivative of PDF f'_ϵ are given by

$$F_\epsilon(x) = \frac{1}{1+e^{-x}}, \quad f_\epsilon(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \quad \text{and} \quad f'_\epsilon(x) = \frac{e^{-x}(e^{-x}-1)}{(1+e^{-x})^3}.$$

To lower bound the variance, we first provide a lower bound of the variance of the adoption state $Y_i(t)$ for each agent $i \in V$ and $t \ge 1$. Since $\mathrm{Var}(Y_i(t)) = q_i(t)(1-q_i(t))$ where $q_i(t) = \mathbb{E}[1 - F_\epsilon(-v - \beta\frac{1}{d_i}\sum_{j\in\mathcal{N}_i} Y_j(t-1))]$, we can derive that, when the intrinsic value of all agents $v_i = v = -\beta - \xi$,

$$\mathrm{Var}\left(Y_i(t)\right) \;\ge\; \left[1 - F_\epsilon(-v)\right]\cdot F_\epsilon(-v-\beta) \;=\; \left[1 - F_\epsilon(\xi+\beta)\right]\cdot F_\epsilon(\xi). \tag{40}$$

Then, we will establish the lower bound on the covariance matrix of adoption states, which is similar to the proof of Lemma 1. Specifically, we show that $\mathrm{Cov}_{\mathbf{Y}(t-1)}\left(\mathbb{E}_{\boldsymbol{\epsilon}(t)}\left[Y_i(t) \mid \mathbf{Y}(t-1)\right],\ \mathbb{E}_{\boldsymbol{\epsilon}(t)}\left[Y_{i'}(t) \mid \mathbf{Y}(t-1)\right]\right)$ can be lower bounded recursively as

$$\begin{aligned}
&\mathrm{Cov}_{\mathbf{Y}(t-1)}\left(\mathbb{E}_{\boldsymbol{\epsilon}(t)}\left[Y_i(t) \mid \mathbf{Y}(t-1)\right],\ \mathbb{E}_{\boldsymbol{\epsilon}(t)}\left[Y_{i'}(t) \mid \mathbf{Y}(t-1)\right]\right)\\
=\ &\mathrm{Cov}\left(1-F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i}Y_j(t-1)}{d_i}\right),\ 1-F_\epsilon\left(-v_{i'}-\beta\frac{\sum_{j'\in\mathcal{N}_{i'}}Y_{j'}(t-1)}{d_{i'}}\right)\right)\\
\geq\ &\frac{(\ell\beta)^2}{d_i d_{i'}}\mathrm{Cov}\left(\sum_{j\in\mathcal{N}_i}Y_j(t-1),\ \sum_{j'\in\mathcal{N}_{i'}}Y_{j'}(t-1)\right) &\text{(41a)}\\
=\ &\frac{\rho_\ell^2}{d_i d_{i'}}\sum_{j\in\mathcal{N}_i}\sum_{j'\in\mathcal{N}_{i'}}\mathrm{Cov}\left(Y_j(t-1),\ Y_{j'}(t-1)\right). &\text{(41b)}
\end{aligned}$$

where (41a) follows from Corollary 4 and (41b) follows from the definition of covariance of linear combinations and the fact that $\rho_\ell = \ell\beta$.

When $t=1$, the covariance matrix can be lower bounded by $c\mathbf{I}$. Therefore, we can lower bound the covariance matrix $\boldsymbol{\Sigma}(t)$ iteratively as

$$\boldsymbol{\Sigma}(t)\geq c\mathbf{I}+\rho_\ell^2\tilde{\mathbf{A}}\boldsymbol{\Sigma}(t-1)\tilde{A}^\top\geq c\mathbf{I}+\rho_\ell^2\tilde{\mathbf{A}}\left[c\mathbf{I}+\rho_\ell^2\tilde{\mathbf{A}}\boldsymbol{\Sigma}(t-2)\tilde{\mathbf{A}}^\top\right]\tilde{\mathbf{A}}^\top\geq\ldots\geq c\left[\mathbf{I}+\sum_{\tau=1}^{t-1}\rho_\ell^{2\tau}\tilde{\mathbf{A}}^\tau\left(\tilde{\mathbf{A}}^\top\right)^\tau\right],$$

where the inequalities follow since (40) and (41).

Furthermore, we have

$$\boldsymbol{\kappa}(t)\geq c\left[\sum_{\tau=1}^{t}\rho_\ell^{2\tau-2}\tilde{\mathbf{A}}^\tau\odot\left(\tilde{\mathbf{A}}^\top\right)^\tau\right]\mathbf{e}.$$

We then bound the difference between $\mathbf{q}(t)$ and $\mathbf{h}\big(\mathbf{q}(t-1)\big)$. Similar to proof of Lemma 4, we obtain

$$\begin{aligned}
\mathbb{E}\left[F_\epsilon\left(X_i\right)-F_\epsilon\left(\nu_i\right)\right]\ &=\ \frac{1}{2}\mathbb{E}\left[f_\epsilon'(\tilde{X}_i)(X_i-\nu_i)^2\right]\ \leq\ -\frac{\ell_f}{2}\mathrm{Var}\left(X_i\right)\ =\ -\frac{\ell_f\beta^2}{2}\mathrm{Var}\left(\frac{1}{d}\sum_{j\in\mathcal{N}_i}Y_j(t-1)\right)\\
&\leq\ -\frac{\ell_f\beta^2 c}{2}\left[\sum_{\tau=1}^{t-1}\rho_\ell^{2\tau-2}\tilde{\mathbf{A}}^\tau\odot\left(\tilde{\mathbf{A}}^\top\right)^\tau\right]\mathbf{e},
\end{aligned}$$

where the first inequality follows because $\tilde{X}_i\in[\xi,\xi+\beta]$ and $f_\epsilon'(x)<-\ell_f$ for x in this range. Therefore,

$$\begin{aligned}
q_i(t)=1-\mathbb{E}[F_\epsilon\left(X_i\right)]\ &\geq\ 1-F_\epsilon\left(\nu_i\right)+\frac{\ell_f\beta^2 c}{2}\left[\sum_{\tau=1}^{t}\rho_\ell^{2\tau-2}\tilde{\mathbf{A}}^\tau\odot\left(\tilde{\mathbf{A}}^\top\right)^\tau\right]\mathbf{e}\\
&=\ h(\mathbf{q}(t-1))_i+\frac{\ell_f\beta^2 c}{2}\left[\sum_{\tau=1}^{t}\rho_\ell^{2\tau-2}\tilde{\mathbf{A}}^\tau\odot\left(\tilde{\mathbf{A}}^\top\right)^\tau\right]\mathbf{e}.
\end{aligned}$$

In the vector form, we have

$$\mathbf{q}(t)\ \geq\ \mathbf{h}(\mathbf{q}(t-1))+\frac{\ell_f\beta^2 c}{2}\left[\sum_{\tau=1}^{t}\rho_\ell^{2\tau-2}\tilde{\mathbf{A}}^\tau\odot\left(\tilde{\mathbf{A}}^\top\right)^\tau\right]\mathbf{e}. \tag{42}$$

Finally, we lower bound the approximation error in a way analogous to Theorem 1. Let $\boldsymbol{\eta}(t) = \frac{\ell_f \beta^2 c}{2}\left[\sum_{\tau=1}^{t} \rho_\ell^{2\tau-2} \tilde{\mathbf{A}}^\tau \odot \left(\tilde{\mathbf{A}}^\top\right)^\tau\right]\mathbf{e}$. Let

$$\overline{\overline{\mu}}_i(t) = \begin{cases} q_i(0) & t=0 \\ 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \overline{\overline{\mu}}_i(t-1)}{d_i}\right) + \eta(t-1)_i & t>0 \end{cases}, \quad \text{for all } i \in V.$$

We will show $\mathbf{q}(t) \geq \overline{\overline{\boldsymbol{\mu}}}(t) \geq \boldsymbol{\mu}(t)$ by induction.

Base case $t=0$: By definition, we have $\mathbf{q}(0) = \overline{\overline{\boldsymbol{\mu}}}(0) = \boldsymbol{\mu}(0)$.

To Show $t=s+1$: Assume that $\mathbf{q}(s) \geq \overline{\overline{\boldsymbol{\mu}}}(s) \geq \boldsymbol{\mu}(s)$. We have

$$\mathbf{q}(s+1) \;\geq\; \mathbf{h}_{\boldsymbol{\delta}(s)}(\mathbf{q}(s)) \;\geq\; \mathbf{h}_{\boldsymbol{\delta}(s)}(\overline{\overline{\boldsymbol{\mu}}}(s)) \;=\; \overline{\overline{\boldsymbol{\mu}}}(s+1) \;\geq\; \mathbf{h}_{\boldsymbol{\delta}(s)}(\boldsymbol{\mu}(s)) \;\geq\; \mathbf{h}(\boldsymbol{\mu}(s)) = \boldsymbol{\mu}(s+1),$$

where the first inequality follows from (42), the second and third inequalities follow Proposition 2 and the induction hypothesis, and the last inequality is trivial because $\boldsymbol{\eta}(s) > 0$.

While the dynamical system $\{\overline{\overline{\boldsymbol{\mu}}}(t)\}$ is not a fixed-point iteration, we can show that it still converges. Let $\boldsymbol{\eta} = \lim_{t\to\infty} \boldsymbol{\eta}(t) = \frac{\ell_f \beta^2 c}{2} \mathcal{B}(G;\rho_\ell)$. By Lemma 10, we can show that $\{\overline{\overline{\boldsymbol{\mu}}}(t)\}$ converges to $\boldsymbol{\nu}^*$, where $\boldsymbol{\nu}^*$ is the fixed-point solution of $\mathbf{h}_{\boldsymbol{\eta}}(\boldsymbol{\nu}) = \boldsymbol{\nu}$. Thus, the following result holds,

$$\mathbf{q}^* \;\geq\; \overline{\overline{\boldsymbol{\mu}}}^* \;=\; \boldsymbol{\nu}^* \;\geq\; \boldsymbol{\mu}^*, \tag{43}$$

where $\overline{\overline{\boldsymbol{\mu}}}^*$ is the limit of $\overline{\overline{\boldsymbol{\mu}}}(t)$. Therefore, we know that $\mathbf{q}^* - \boldsymbol{\mu}^*$ will be lower bounded by $\overline{\overline{\boldsymbol{\mu}}}^* - \boldsymbol{\mu}^*$. We have

$$\overline{\overline{\boldsymbol{\mu}}}^* - \boldsymbol{\mu}^* \;=\; \mathbf{h}(\overline{\overline{\boldsymbol{\mu}}}^*) - \mathbf{h}(\boldsymbol{\mu}^*) + \boldsymbol{\eta} \;\geq\; \rho_\ell \tilde{\mathbf{A}}\left(\overline{\overline{\boldsymbol{\mu}}}^* - \boldsymbol{\mu}^*\right) + \boldsymbol{\eta},$$

where the inequality follow from Assumption 3.

As a result, we can arrive at the lower bound

$$\overline{\overline{\boldsymbol{\mu}}}^* - \boldsymbol{\mu}^* \geq \left(\mathbf{I} - \rho_\ell \tilde{\mathbf{A}}\right)^{-1} \boldsymbol{\eta} = \frac{\ell_f \beta^2 c}{2}\left(\mathbf{I} - \rho_\ell \tilde{\mathbf{A}}\right)^{-1} \mathcal{B}(G;\rho_\ell),$$

and conclude the proof.

□

Appendix E: Supplementary Numerical Experiments on the FPA Scheme

E.1. Illustration of the 10-Node Example Instance

To offer a clear illustration of this instance, we construct an undirected network comprising 10 nodes. The network structure is visualized in Fig. 1a, while the intrinsic values assigned to each agent are detailed in Table 3. We set the network effect intensity at $\beta = 3.5$ and assume that the random noise distribution is $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0,1)$ for all $i \in V$ and $t \geq 0$. The characteristics of this example network, along with numerical results obtained from different models, are presented in Table 3.

Table 3 Characteristics and results of the 10-node example instance

Node	In-degree d	Intrinsic value v	q^*	μ^*	FPA error	q^{MM}	MM error
0	5	-1.7064	0.5126	0.5292	0.0166	0.1536	-0.3590
1	7	-1.2453	0.5932	0.6069	0.0137	0.2235	-0.3697
2	4	-0.8789	0.6325	0.6524	0.0199	0.2934	-0.3391
3	4	-3.9454	0.1442	0.1221	-0.0222	0.0190	-0.1253
4	3	-0.0822	0.7827	0.8219	0.0393	0.4795	-0.3032
5	5	-3.4441	0.1933	0.1731	-0.0202	0.0309	-0.1624
6	3	-0.2877	0.7341	0.7755	0.0414	0.4286	-0.3055
7	2	-2.9084	0.3287	0.2849	-0.0438	0.0517	-0.2770
8	2	-1.2859	0.6702	0.7646	0.0944	0.2166	-0.4536
9	1	-0.6963	0.7416	0.8786	0.1371	0.3326	-0.4090

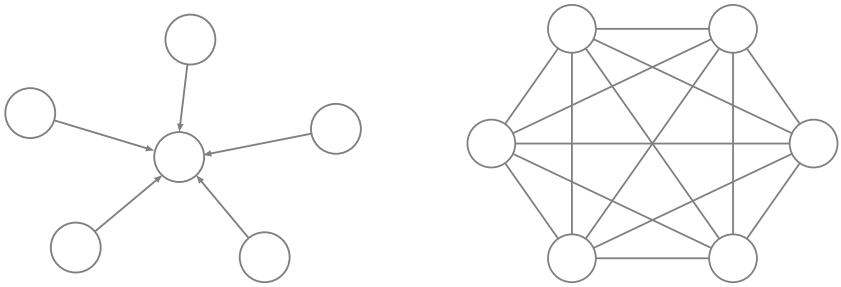
Note: p^* is calculated by first constructing a 1,024-state MC according to Section 2.2 and calculating the stationary distribution. μ^* is calculated by conducting fixed-point iteration according to (5), and FPA error equals $(\mu_i^* - q_i^*)$. p^{MM} is calculated as $\mathbb{E}[\mathbb{1}\{v_i + \epsilon_i \geq 0\}]$, and MM error equals $(\mu_i^{\mathsf{MM}} - q_i^*)$.

E.2. Numerical Experiments on Highly-Structured Symmetric Networks

To illustrate the exact performance of the FPA scheme, we focus on two kinds of highly-structured symmetric networks, namely directed star networks and complete networks. These simple and symmetric structures make it easier to calculate the limiting adoption probability. We further simplify the diffusion instance by setting the intrinsic value of all agents in the network to be the same as v. This allows us either to directly compute the limiting adoption probability or to construct an MC with a much smaller state space.

Network instances:

- *Directed star networks.* A star network consists of a central node and several surrounding nodes. We consider the directed version where the edges only point from surrounding nodes to the central node, not vice versa. Figure 10a shows an example of network size $n = 6$.
- *Complete networks.* A complete network is the network where each node is directly connected to every other node. Figure 10b shows an example of network size $n = 6$.



(a) Directed star network (b) Complete network

Figure 10 Illustration of highly structured symmetric network structure.

For directed star networks, the adoption decisions of surrounding nodes are independent of each other. Therefore, we can directly calculate the limiting adoption probability of the central node as

$$q = \sum_{i=0}^{n-1} \binom{n-1}{i} \left(1 - F_\epsilon(-v)\right)^i F_\epsilon(-v)^{n-1-i} \cdot \left[1 - F_\epsilon\left(-v - \beta \frac{i}{n-1}\right)\right].$$

For complete networks, we can construct a more efficient MC by using the number of adopted agents as the MC states, rather than considering the combination of all agents' adoption states. The transition probability of this MC can be defined as

$$P(i,j) = \sum_{k=0}^{\min\{i,j\}} \binom{i}{k} \left[1 - F_\epsilon\left(-v - \beta \frac{i-1}{n-1}\right)\right]^k F_\epsilon\left(-v - \beta \frac{i-1}{n-1}\right)^{i-k} \binom{n-i}{j-k} \cdot \left[1 - F_\epsilon\left(-v - \beta \frac{i}{n-1}\right)\right]^{j-k} F_\epsilon\left(-v - \beta \frac{i}{n-1}\right)^{n-i-j+k}.$$

Hence, the limiting adoption probability for this MC can be directly calculated.

We measure FPA performance by the percentage error (PE) of the representative node, given in the following equation:

$$\text{PE} = \frac{\mu_i^* - q_i^*}{q_i^*} \cdot 100\%.$$

In directed star networks, we focus solely on the central node because the surrounding nodes have zero in-degree and can thus be perfectly approximated by the FPA scheme. In complete networks, the PE is identical for all nodes. Therefore, the PE for any arbitrary node in a complete network is equivalent to the mean average percentage error.

To assess FPA performance, we investigate two scenarios for both types of network structures: (i) a sequence of diffusion instances with different intrinsic values, and (ii) a sequence of diffusion instances with different network sizes. For these experiments, we set the network effect intensity to be $\beta = 1$ and generate the random noise $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0,1)$.

(i) *The accuracy with regard to intrinsic values.* We choose the intrinsic value v from -5 to 5 in increments of 0.1. These instances are tested on networks of size $n \in \{10, 20, 30\}$. Figure 11 shows the PE of both network structures at different intrinsic values. Overall, all instances have a small absolute percentage error (less than

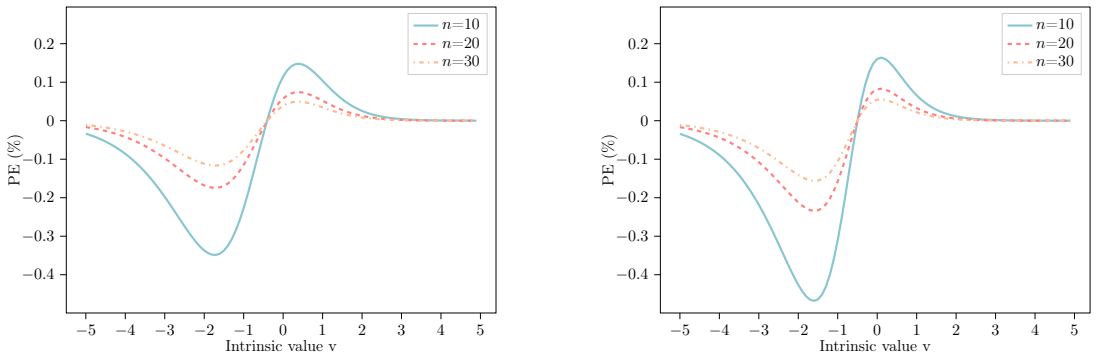


Figure 11 PE versus intrinsic value. Left: Directed star network; Right: Complete network.

0.5%), illustrating the high accuracy of the FPA solution. We notice that the PE curves of different network

structures possess similar shapes, however, the exact values are slightly different. In general, the FPA scheme tends to underestimate the adoption probability when the intrinsic values are small and overestimate it when they are large. There exist two critical points at around $v = -1.7$ and $v = 0.4$ where the PE reaches extremes. These points exhibit the worst cases and align with regions where the CDF $F_\epsilon(\cdot)$ has the largest curvature.

(ii) *The accuracy with regard to network size.* We then focus on instances with intrinsic values at the two previously mentioned critical points $v \in \{-1.7, 0.4\}$. We choose the network size n from 2 to 50. Figure 12 shows the PE across these different network sizes. Regardless of the network structure and the intrinsic

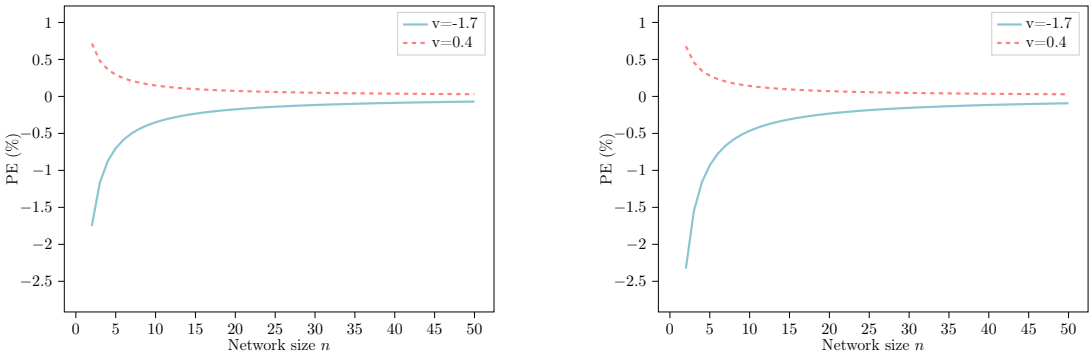


Figure 12 **PE versus network size. Left: Directed star network; Right: Complete network.**

values, PE converges to 0 rapidly when the network size increases. This can be theoretically confirmed, and we have explicitly demonstrated it in Corollary 1 or a refined version in Theorem 2. Our findings show that, for highly-structured networks, the FPA scheme offers excellent approximation quality and exhibits asymptotic convergence as the network size grows.

E.3. Agent-Based Simulation Settings

In this section, we discuss the simulation methods used in our work, which serve as benchmarks for comparing the FPA scheme. First, we describe the naïve ABS method, which is used both for benchmarking and establishing the underlying ground truth. We also empirically examine the warm-up period for the simulation. Then, we will introduce the more advanced A-ABS method.

E.3.1. Naïve ABS Method. The simulation starts with all agents in a non-adopted state. At each time step, the state of each agent is updated based on the previous time step using the agent behavior model. Specifically, for agent i at time t, we determine their adoption state based on $\mathbf{Y}(t-1)$ using the following equation:

$$Y_i(t) = \mathbb{1}\left\{v_i + \beta \frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i} + \epsilon_i(t) \geq 0\right\}.$$

We designate the first 1,000 time steps as the warm-up period to allow the system to reach a steady state. These initial steps are discarded from the analysis to avoid transient bias. After the warm-up period, we collect T time steps of the adoption states. Following the first equation in (3), the adoption probability for agent i is estimated as:

$$\hat{q}_i = \frac{1}{T}\sum_{t=1}^{T} Y_i(t).$$

For the ground truth, we set the simulation run length to 100,000 steps beyond the warm-up period. When testing the performance of ABS methods, we stop the simulation once the real-time MAPE of ABS methods falls below that of the FPA scheme.

Additionally, we conduct experiments to empirically demonstrate why the warm-up period can be set as 1,000 time steps. Rather than focusing on individual states of the MC, we track the average cumulative adoption proportion across the entire population as an indicator, which is given by:

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n}\sum_{i\in V}Y_i(t).$$

In Figure 13, we show how the average cumulative adoption proportion evolves over time. We test on 4 different diffusion instances, each represented by a randomly sampled Erdős-Rényi network $G(n,p(n))$. We choose the network size from $n\in\{10,100,1000,10000\}$ and keep the probability of edge existence to be $p(n)=0.1$. We set the network effect intensity to be $\beta=1$ and generate the random noise $\epsilon_i(t)\overset{\text{i.i.d.}}{\sim}\text{Logistic}(0,1)$. Our results show that after 1,000 time steps, all tested trajectories reach a steady state. Larger networks

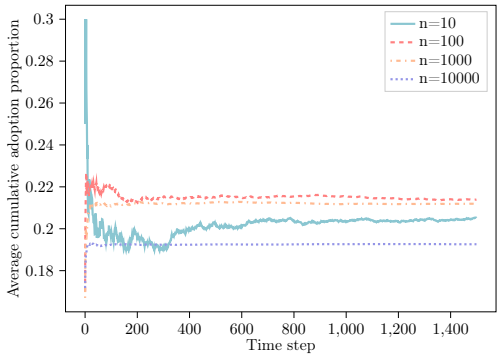


Figure 13 Average cumulative adoption proportion versus the time steps of Naïve ABS

tend to reach a steady state more quickly. Additional tests with varying parameters yielded similar results, leading us to conclude that a warm-up period of 1,000 time steps is sufficient for our context.

E.3.2. Accelerated ABS method. The Naïve ABS method is computationally expensive due to the large number of time steps required to move past the transient phase and the need to sample many adoption states for accurate estimation. Furthermore, because the adoption state $Y_i(t)$ is correlated over time and across the network, the time required to achieve a target accuracy becomes even more substantial.

We observe that the adoption state $Y_i(t)$ is a compound Bernoulli variable, with uncertainty in estimating the limiting adoption probability $\mathbf{q}^*$ arising from two sources: (i) the neighbors' adoption states $\mathbf{Y}(t-1)$ and (ii) the random noise $\boldsymbol{\epsilon}(t)$. The Naïve ABS method incorporates both sources of uncertainty, which require large samples for an accurate estimation.

Inspired by the second equation in (3), we leverage the analytical form of the adoption probability to improve sampling efficiency. In particular, we reduce one source of uncertainty. Given $\mathbf{Y}(t-1)$, we can express the adoption probability $q_i(t)$ as follows:

$$q_i(t)=\mathbb{E}_{\mathbf{Y}(t-1)}\mathbb{E}_{\boldsymbol{\epsilon}(t)}[Y_i(t)|\mathbf{Y}(t-1)]=\mathbb{E}_{\mathbf{Y}(t-1)}\left[\mathbb{E}_{\boldsymbol{\epsilon}(t)}\left[\mathbb{1}\left\{v_i+\beta\frac{\sum_{j\in\mathcal{N}_i}Y_j(t-1)}{d_i}+\epsilon_i(t)\geq 0\right\}\middle|\mathbf{Y}(t-1)\right]\right]$$

$$= \mathbb{E}_{\mathbf{Y}(t-1)}\left[1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i}\right)\right].$$

Since $\mathbf{Y}(t-1)$ is observable during the simulation, we can estimate the adoption probability more efficiently as follows:

$$\hat{q}_i = \frac{1}{T}\sum_{t=1}^{T}\left[1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{n_i}\right)\right].$$

This A-ABS approach reduces the computational burden by replacing the need for large samples of adoption states with conditional probability evaluations, improving the efficiency of the simulation process.

E.4. Numerical Experiments for Power-Law Networks

We extend our examination of the FPA scheme to another important class of random networks, i.e., power-law networks. This class of networks exhibits a degree distribution that follows a power-law pattern. We consider a sequence of directed power-law networks with n nodes and define the associated CDF of the degree distribution as $F_d(\cdot\,; n)$. The network in-degrees and out-degrees are generated using the following CDF:

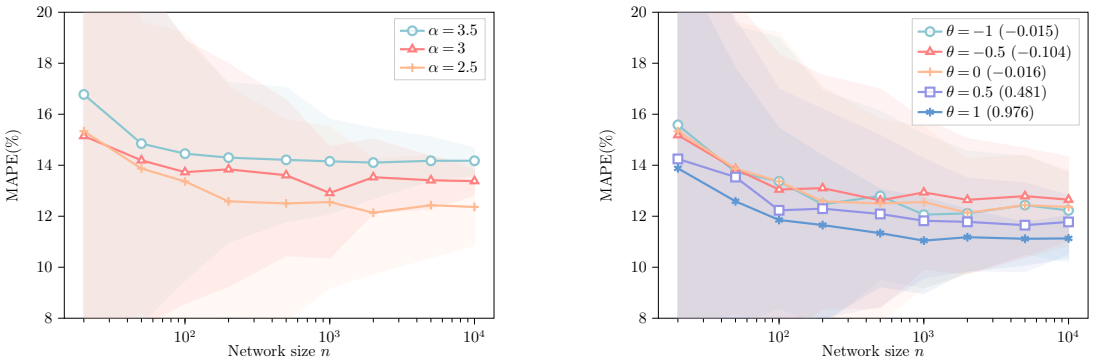
$$F_d(x;n) \;=\; \mathbb{P}(d \le x) \;=\; \frac{1 - \left(\frac{x}{d_{\min}}\right)^{1-\alpha}}{1 - \left(\frac{d_{\max}}{d_{\min}}\right)^{1-\alpha}} \quad \text{for} \quad d_{\min} \le x \le d_{\max} \;=\; n,$$

where α is the exponent of power-law distribution. We set $d_{\min}$ to be 2 and $d_{\max}$ to be n. Correspondingly, the probability mass function satisfies $f_d(x) \propto x^{-\alpha}$, which aligns with the conventional definition of a power-law distribution. Power-law networks often pose significant challenges for the analysis and optimization on networks due to the prevalence of low-degree nodes in such networks. Focusing on power-law networks with $d_{\min} = 2$ and $\rho = 0.875$ (see Section 5) allows us to test the limit of the FPA scheme.

We generate power-law network based on $F_d(\cdot\,; n)$ following the approach proposed by Huang et al. (2022). Detailed information on the generation process is included in Appendix E.5 for completeness. In this generation process we use an auxiliary parameter θ to account for the pairwise correlation between the in-degree and out-degree sequences. Although θ is not the exact correlation between these two sequences, it approximates the actual correlation between in-degrees and out-degrees, particularly for large values of n.

We conduct two sets of experiments to test the FPA scheme across different power-law exponents α and pairwise correlations θ. For each parameter combination, we conduct 100 repetitions to ensure stable performance metrics. The results are presented in Figure 14. In general, The FPA scheme still performs reasonably well. Also, we observe a consistent decrease in the MAPE as n increases across all tested parameter pairs, albeit at a relatively modest pace in comparison with Erdős-Rényi networks. Additionally, power-law networks exhibit both a higher mean and greater variance in MAPE. The increased mean MAPE in power-law networks is largely attributable to a higher proportion of low in-degree nodes. The increased variance, on the other hand, is primarily due to the more intricate structural variations inherent to power-law networks when specified parameters are used.

We note that the power-law exponent α has a crucial impact on the degree distribution. Typically, power-law networks feature an $\alpha > 2$ to avoid divergence in the expected degree. In the case where $\alpha = 3$, the network adheres to a model generated through the preferential attachment process. Accordingly, we select



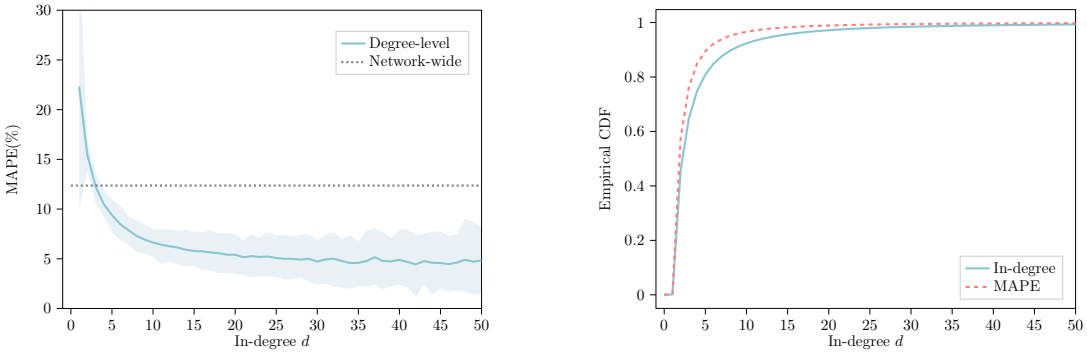
(a) MAPE against α (b) MAPE against θ

Figure 14 FPA Performance on power-law networks for different α and θ values. All horizontal axes are in the log scale. Shaded areas represent the 95% confidence interval. In the legend of the right subfigure, numbers within parentheses represent empirical correlations between in-degrees and out-degrees.

α from the set $\{2.5, 3, 3.5\}$ and set $\theta = 0$ to construct Figure 14a. It is evident that as α increases, the proportion of low-degree nodes also increases. Consistent with our theoretical analysis, we observe that the MAPE tends to rise as α increases.

In the second experiment, we generate the in-degree and out-degree sequences with $\theta \in \{-1, -0.5, 0, 0.5, 1\}$. Corollary 2 shows that the FPA scheme's performance is related to the imbalance level of the network, which can be captured by this pairwise correlation coefficient θ. Specifically, a large positive θ indicates a strong positive correlation between in-degree and out-degree sequences, resulting in a more balanced network. Conversely, a negative θ, suggests a more imbalanced network. From Figure 14b, the MAPE remains relatively stable when θ ranges between -1 and 0. However, it substantially diminishes as θ becomes positive, which aligns with our theoretical findings that the FPA scheme performs worse on more imbalanced networks.

Finally, we focus on FPA performance for nodes with different in-degrees d. For illustration, we choose instances with $\alpha = 2.5$ and $\theta = 0$. In Figure 15a, we illustrate how the MAPE varies with respect to the in-degree d. We can notice the MAPE consistently decreases for nodes with $d \geq 1$ as d increases. For nodes



(a) MAPE against d (b) Empirical CDF for in-degree and MAPE

Figure 15 FPA Performance on power-law networks with regard to the in-degree values. The dotted line in the left subfigure is the average network-wide MAPE. Shaded area represents 95% confidence interval.

with more than 10 in-neighbors, the MAPE diminishes to less than 6.5%. Furthermore, the network-wide MAPE stands at approximately 12.36%; notably, only nodes with an in-degree of less than 3 exhibit errors above this level. In Figure 15b, we extend our analysis by displaying the empirical CDF for the in-degree distribution and the MAPE. Given the nature of power-law networks, a substantial number of nodes exhibit low in-degree. Moreover, we can notice that the empirical CDF of the node-level MAPE is higher than that of in-degree, indicating these low in-degree nodes are also associated with larger errors. Specifically, 84.89% of the total error is attributable to agents with fewer than 5 in-neighbors, and 95.97% of the error can be attributed to agents with fewer than 10 in-neighbors.

E.5. Supplementary Discussions on Random Networks.

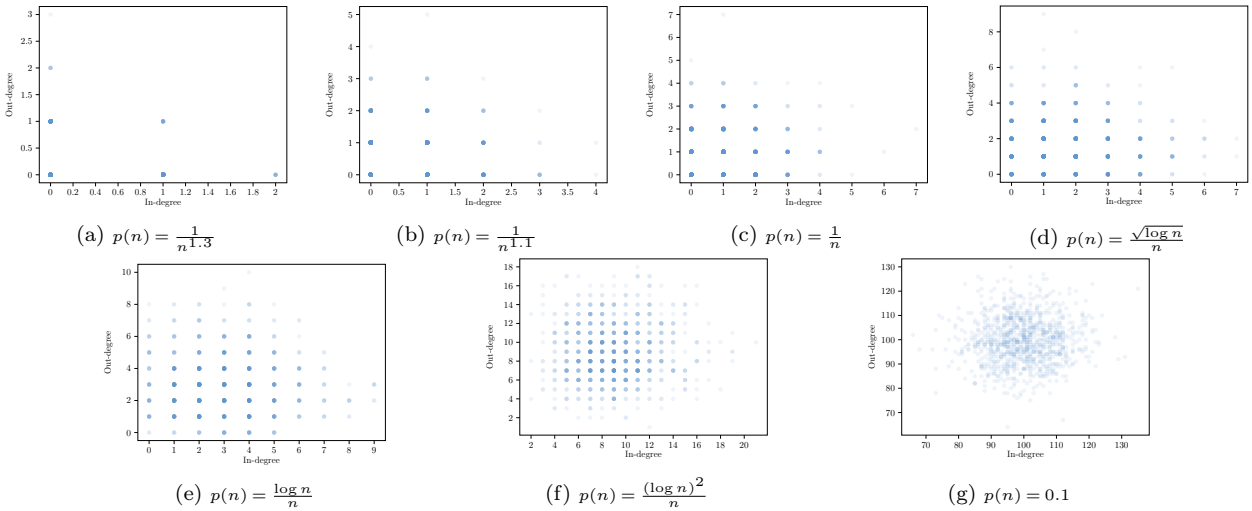
In our numerical experiments of random networks, we generate our data following the setup outlined in Huang et al. (2022), which also offers an excellent discussion on the key properties of this type of network. In the following, we revisit some of the discussions on parameter selection and instance construction for both Erdős-Rényi and power-law networks, supplementing them with additional numerical illustrations for more robust empirical support. For more details, please refer directly to this paper.

(i) *Erdős-Rényi networks*. In the asymptotic analysis of Erdős-Rényi networks, the density $p(n)$ plays a pivotal role in shaping the structural attributes of the network. Some critical cases are outlined as follows:

- When $p(n) = o(n^{-2})$, the Erdős-Rényi networks are empty almost surely (Erdős et al. 1960).
- When $p(n) = \mathcal{O}(n^{-(1+\epsilon)})$ for some $\epsilon > 0$, the expected in-degree and out-degree vanishes asymptotically. Such networks are called *very sparse* networks. They are probabilistically acyclic and fragmented.
- When $p(n) = \Theta(n^{-1})$, the expected in-degree and out-degree remain asymptotically bounded and positive. Such networks are called *critically sparse* networks. At this point, a phase transition occurs: as $p(n)$ increases from $\frac{1}{n} - \mathcal{O}(n^{-\frac{4}{3}})$ to $\frac{1}{n} + O(n^{-\frac{4}{3}})$, smaller components merge into a giant component comprising a positive fraction of nodes, and cycles begin to form (Janson et al. 1993).
- When $p(n) = \omega(\frac{\log n}{n})$, networks are called *dense* networks. These networks are highly likely to be connected, contain many cycles, and are asymptotically regular and balanced. Both in-degree and out-degree distributions concentrate around the mean value and converge asymptotically to a normal distribution.

It is important to note that dense networks are asymptotically regular and balanced. This underlying property aligns our numerical findings with the theoretical implications with regard to the imbalance level of networks. In Figure 16, we illustrate the distribution of in-degree and out-degree pairs for each node in the Erdős-Rényi networks with different densities. As evidenced by Figure 16, the distribution of in-degrees and out-degrees approaches a normal distribution as $p(n)$ increases. More importantly, this trend also shows that the in-degree and the out-degree of a specific node become close to each other, contributing to a more balanced network structure.

(ii) *Power-law networks*. To avoid notation confusion, in this part, we let d_i^{in} and d_i^{out} denote the in-degree and out-degree of node i. To construct a power-law network, we require both the in-degrees $d_1^{\text{in}}, d_2^{\text{in}}, \ldots, d_n^{\text{in}}$ and the out-degrees $d_1^{\text{out}}, d_2^{\text{out}}, \ldots, d_n^{\text{out}}$ are i.i.d. sampled from distribution $F_d(\cdot, n)$. One distinctive aspect of



(a) $p(n) = \frac{1}{n^{1.3}}$ (b) $p(n) = \frac{1}{n^{1.1}}$ (c) $p(n) = \frac{1}{n}$ (d) $p(n) = \frac{\sqrt{\log n}}{n}$

(e) $p(n) = \frac{\log n}{n}$ (f) $p(n) = \frac{(\log n)^2}{n}$ (g) $p(n) = 0.1$

Figure 16 **In-degree and out-degree distributions of Erdős-Rényi networks with different densities $p(n)$.**

our experiments with power-law networks is the introduction of a pairwise correlation parameter to capture the imbalance level of the network. A valid correlated in-degree sequence and out-degree sequence can be generated using the following procedure:

- Sample i.i.d. in-degrees $d_1^{\text{in}}, d_2^{\text{in}}, \ldots, d_n^{\text{in}}$ from the power-law distribution. Without loss of generality, assume this sequence is sorted in descending order.
- Sample i.i.d. random variables $Z_1, Z_2, \ldots, Z_n$ as follows: for each $i \in V$, $Z_i = 1$ with probability $|\theta|$, and $Z_i = 0$ with probability $1 - |\theta|$, where $\theta \in [-1, 1]$ is the parameter used to control the correlation. This parameter θ is not necessarily the correlation $\text{Cov}(d_i^{\text{in}}, d_i^{\text{out}})$.
- Define sets of nodes $I_0 : \{i : Z_i = 0, 1 \le i \le n\}$ and $I_1 : \{i : Z_i = 1, 1 \le i \le n\}$.
- If $\theta \ge 0$, set $d_i^{\text{out}} = d_i^{\text{in}}$ for $i \in I_1$ and set $\{d_i^{\text{out}} : i \in I_0\}$ by a random permutation of $\{d_i^{\text{in}} : i \in I_0\}$; If $\theta < 0$, set $d_i^{\text{out}} = d_{n-i+1}^{\text{in}}$ for $i \in I_0$ and set $\{d_i^{\text{out}} : i \in I_1\}$ by a random permutation of $\{d_{n-i+1}^{\text{in}} : i \in I_1\}$.
- Use a configuration model (Molloy and Reed 1995, Newman et al. 2001) to construct the directed random network with given in- and out-degree sequences.

Under this construction, when $\theta \ge 0$, the correlation is $\theta + O(n^{-1})$, so it asymptotically equals θ. When $\theta < 0$, the generated pairwise correlation may deviate from θ, and different values of θ yield similar degree sequences, as evidenced by Figure 17.

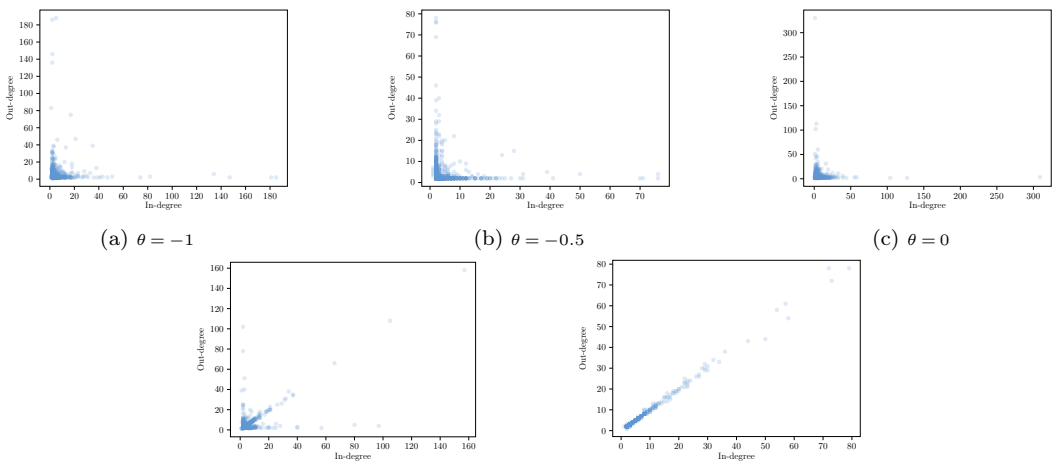
Appendix F: Supporting Arguments for Section 6

F.1. Supporting Arguments for Section 6.1

We first provide the proof to characterize the optimality gap of the approximate IM problem.

Proof of Proposition 3: The proof largely follows Corollary 2, where we bound the scaled ℓ_1-norm of the FPA error. Therefore, the regret can be bounded by

$$\text{Regret}(S^{\text{FPA}}) = \sum_{i \in V} q_i^*(S^*) - \sum_{i \in V} q_i^*\left(S^{\text{FPA}}\right)$$



(a) $\theta = -1$ (b) $\theta = -0.5$ (c) $\theta = 0$

(d) $\theta = 0.5$ (e) $\theta = 1$

Figure 17 **In-degree and out-degree distributions of power-law networks with different pairwise correlations θ.**

$$
\begin{aligned}
&= \sum_{i\in V} q_i^*\left(S^*\right) - \sum_{i\in V} \mu_i^*\left(S^*\right) + \sum_{i\in V} \mu_i^*\left(S^*\right) - \sum_{i\in V} \mu_i^*\left(S^{\mathrm{FPA}}\right) + \sum_{i\in V} \mu_i^*\left(S^{\mathrm{FPA}}\right) - \sum_{i\in V} q_i^*\left(S^{\mathrm{FPA}}\right) \\
&\le \left\|\mathbf{q}^*\left(S^*\right) - \boldsymbol{\mu}^*\left(S^*\right)\right\|_1 + \left(\sum_{i\in V} \mu_i^*\left(S^*\right) - \sum_{i\in V} \mu_i^*\left(S^{\mathrm{FPA}}\right)\right) + \left\|\mathbf{q}^*\left(S^{\mathrm{FPA}}\right) - \boldsymbol{\mu}^*\left(S^{\mathrm{FPA}}\right)\right\|_1 \\
&\le \left\|\mathbf{q}^*\left(S^*\right) - \boldsymbol{\mu}^*\left(S^*\right)\right\|_1 + \left\|\mathbf{q}^*\left(S^{\mathrm{FPA}}\right) - \boldsymbol{\mu}^*\left(S^{\mathrm{FPA}}\right)\right\|_1 \\
&\le 2C_\rho\sqrt{n\|\mathcal{C}(G;\rho)\|_1},
\end{aligned}
$$

where the first inequality holds trivially, the second inequality follows from the optimality of S^{FPA} for approximate IM problem (20), and the third inequality follows from Corollary 2. □

In the next, we discuss the applicability of Assumption 4.

Instances that satisfy Assumption 4. As we mentioned immediately after the assumption, the nonprogressive LT model is a specific instance that meets this assumption. Recall we can recover the LT model by setting $v_i = -0.5$ and $\epsilon_i(t) \sim \mathcal{U}(-0.5, 0.5)$ for all $i \in V$ and $t \ge 1$. Therefore, for any $\beta > 0$, CDF F_ϵ can be expressed as $F_\epsilon(x) = \mathbb{1}\{x \ge -0.5\} \cdot (x + 0.5)$ on range $[0.5 - \beta, 0.5]$, which is convex. Additionally, some other diffusion instances related to common utility models can also meet Assumption 4. Some examples are: (i) Linear probability model: $v_i \ge -c$ and $\epsilon_i(t) \sim \mathcal{U}(-c, c)$ for all $i \in V$, $t \ge 1$. (ii) Logit model: $v_i \ge 0$ and $\epsilon_i(t) \sim \text{Logistic}(0, s)$ and for all $i \in V$, $t \ge 1$. (iii) Probit model: $v_i \ge 0$ and $\epsilon_i(t) \sim \mathcal{N}(0, s)$ and for all $i \in V$, $t \ge 1$. For many general distributions, the convexity assumption essentially requires the intrinsic values to be appropriately lower bounded.

Finally, we show the proof of the submodularity for the approximate IM objective.

Proof of Theorem 4: Consider two seed set $S_1 \subseteq S_2 \subseteq V$ and an additional user $w \in V \setminus S_2$, it is sufficient to show that $\boldsymbol{\mu}^*(S_2 + \{w\}) - \boldsymbol{\mu}^*(S_2) \le \boldsymbol{\mu}^*(S_1 + \{w\}) - \boldsymbol{\mu}^*(S_1)$.

We consider constraints (20b) and (20c) as the dynamic system, that is, $\boldsymbol{\mu}(t) = \mathbf{h}(\boldsymbol{\mu}(t-1))$. We can notice that, for different seed sets, the transition function $\mathbf{h}$ is not the same. However, for all the users that are not selected as seed users, the transition function for the corresponding element is the same. With a little abuse of notation, in the following proof, we use $\mathbf{h}$ to denote the transition function for all users in $V \setminus (S_2 + \{w\})$.

We want to show that at any time step $t \geq 1$, the inequality $\boldsymbol{\mu}(S_2+\{w\},t)-\boldsymbol{\mu}(S_2,t) \leq \boldsymbol{\mu}(S_1+\{w\},t)-\boldsymbol{\mu}(S_1,t)$ always holds. For user $i \in S_2$, $\mu_i(S_2+\{w\},t)-\mu_i(S_2,t)=0 \leq \mu_i(S_1+\{w\},t)-\mu_i(S_1,t)$. For user w, $\mu_w(S_2+\{w\},t)-\mu_w(S_2,t)=1-\mu_w(S_2,t) \leq 1-\mu_w(S_1,t) \leq \mu_w(S_1+\{w\},t)-\mu_w(S_1,t)$. The above two inequalities hold because of Proposition 2(i). For all the other users in $V \setminus (S_2 \cup \{w\})$, we show by induction.

Base case $t=0$: First of all, $\boldsymbol{\mu}(S,0)=\mathbf{0}$ for all $S \subseteq S_2 \cup \{w\}$ by definition. Therefore,

$$\boldsymbol{\mu}(S_2+\{w\},0)-\boldsymbol{\mu}(S_2,0)=\boldsymbol{\mu}(S_1+\{w\},0)-\boldsymbol{\mu}(S_1,0).$$

Assume $t=s$: The induction hypothesis holds such that

$$\boldsymbol{\mu}(S_2+\{w\},s)-\boldsymbol{\mu}(S_2,s) \leq \boldsymbol{\mu}(S_1+\{w\},s)-\boldsymbol{\mu}(S_1,s).$$

Show $t=s+1$: We have

$$\begin{aligned}
\boldsymbol{\mu}(S_2+\{w\},s+1)-\boldsymbol{\mu}(S_2,s+1) &= \mathbf{h}(\boldsymbol{\mu}(S_2+\{w\},s))-\mathbf{h}(\boldsymbol{\mu}(S_2,s)) \\
&\leq \mathbf{h}(\boldsymbol{\mu}(S_2,s)+\boldsymbol{\mu}(S_1+\{w\},s)-\boldsymbol{\mu}(S_1,s))-\mathbf{h}(\boldsymbol{\mu}(S_2,s)) \\
&\leq \mathbf{h}(\boldsymbol{\mu}(S_1+\{w\},s))-\mathbf{h}(\boldsymbol{\mu}(S_1,s)) \\
&= \boldsymbol{\mu}(S_1+\{w\},s+1)-\boldsymbol{\mu}(S_1,s+1),
\end{aligned}$$

where the first inequality comes from Proposition 2(i) and the second inequality comes from Assumption 4.

When t tends to infinity, we get the fixed-point solution $\boldsymbol{\mu}^*(S_2+\{w\})-\boldsymbol{\mu}^*(S_2) \leq \boldsymbol{\mu}^*(S_1+\{w\})-\boldsymbol{\mu}^*(S_1)$, and hence the submodularity is proved. □

F.2. Experiments on IM Problem

In the experiments, we consider two scenarios, one satisfies Assumption 4 and thus leads to a submodular influence function, while the other does not. For both scenarios, we assume that the intrinsic value $v_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}(-4,0)$ and $\beta=3.5$. In addition, we assume the random noise to be $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \mathcal{U}(-4,4)$ in the submodular case, while $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0,1)$ in the nonsubmodular case.

The well-known greedy framework selects one user at each iteration which leads to the largest total adoptions. We refer to the algorithm that embeds the FPA solution into this greedy framework for the total influence evaluation as the greedy-FPA algorithm. We randomly generate some small network instances to illustrate that greedy-FPA can find a near-optimal solution. Although there is no theoretical guarantee for the nonsubmodular case, it is interesting to observe from the results that the FPA solutions are still of good quality. For either scenario, we generate 100 diffusion instances with random graph $G(15,0.5)$ and set the number of seed users to 5. We enumerate all the subsets to find the optimal seed set and evaluate the diffusion influence using Naïve ABS method. In Table 4, we show the numerical results of the greedy-FPA algorithm.

We notice that in both the submodular and nonsubmodular scenarios, the greedy-FPA algorithm can generate a near-optimal IM solution and even uncover the exact optimal solution for a large portion of instances. Meanwhile, the greedy-FPA algorithm has a slightly better performance in the submodular case than in the nonsubmodular case but even in the nonsubmodular problem instances, it remains quite practical.

Table 4 Numerical results of greedy-FPA algorithm for IM problem.

Scenario	Percentage of instances where the optimal seeding is recovered	Optimality Gap (%)	
		Mean	Max
Submodular	91	0.0194	0.4704
Non-submodular	84	0.0685	1.7444

Furthermore, we choose a real-world network—*Caltech36* as introduced in Section 5.3 and compare the performance of greedy-FPA with the traditional IM heuristics. Recall that the instance includes 765 agents with an average number of neighbors of 43. We define several benchmark strategies as follows. The DEG and EIG schemes are motivated by the important role of the centrality measures in diffusion discussed in the network economics literature (e.g., Ballester et al. 2006, Jackson 2010). We include them for completeness, but as substantiated in the numerical experiments, by overlooking the idiosyncratic features of the agents, these schemes are dominated by the FPA-based heuristic.

Benchmarks:

- *Greedy and Naïve ABS* (greedy-ABS): This is the classical algorithm used for the IM problem. The Naïve ABS is embedded into a greedy framework for influence evaluation. The length of the Naïve ABS run is set to 100,000 after the warm-up period.
- *Greedy and the FPA solution* (greedy-FPA): This is our proposed algorithm. We embed the FPA solution into a greedy framework for an influence evaluation.
- *Greedy and low-resolution ABS* (greedy-low-ABS): The Naïve ABS is embedded into a greedy framework for an influence evaluation. The length of the Naïve ABS run is set to 50 so that the runtime is at the same scale as that of the FPA scheme.
- *Degree centrality* (DEG): Set K users with the largest degrees to be seed users.
- *Eigenvector centrality* (EIG): Set K users with the largest eigenvector centralities to be seed users.
- *Model misspecification without network effect* (MM): This benchmark considers the misspecified model that ignores the network effect in the IM problem. This is the same as setting K users with the smallest intrinsic value to be seed users.
- *Random* (RAN): Randomly select K users to be seed users.

Figure 18 demonstrates the relative loss of the expected limiting adoptions compared with greedy-ABS against the number of seed users. Similarly, we also consider both the submodular and nonsubmodular cases.

When the number of seed users increases from 0 to 30, the difficulty of the IM problem increases since the number of feasible solutions also increases. We observe that regardless of the number of seed users, the performance of greedy-FPA matches that of greedy-ABS nearly perfectly. It significantly outperforms all the other benchmarks. This is no surprise to us, as it is driven by the high accuracy of the FPA scheme. In particular, we also notice that the performance of the greedy framework with the Naïve ABS method degrades

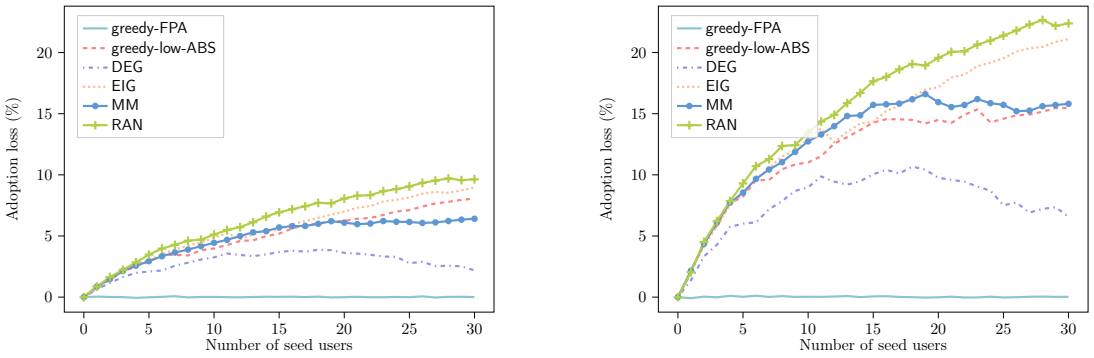


Figure 18 **Performance of different IM algorithms. Left: Submodular case; Right: Non-submodular case.**

drastically when the simulation length of the Naïve ABS procedure is small. Compared with greedy-low-ABS, greedy-FPA achieves an improvement of 8.90% and 18.42% when $K = 30$ in the submodular and nonsubmodular cases. In short, we conclude that, by offering a significant efficiency gain, greedy-FPA outperforms greedy-ABS in solving the IM problem.

F.3. Supporting Arguments for Section 6.2

We first provide the proof to characterize the optimality gap of the approximate pricing problem.

Proof of Proposition 4: The proof aligns with Proposition 3. Therefore, the regret can be bounded by

$$\begin{aligned}
\text{Regret}(\mathbf{p}^{\text{FPA}}) &= \mathbf{q}^*(\mathbf{p}^*)^\top \mathbf{W}\mathbf{p}^* - \mathbf{q}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W}\mathbf{p}^{\text{FPA}} \\
&= \mathbf{q}^*(\mathbf{p}^*)^\top \mathbf{W}\mathbf{p}^* - \boldsymbol{\mu}^*(\mathbf{p}^*)^\top \mathbf{W}\mathbf{p}^* + \boldsymbol{\mu}^*(\mathbf{p}^*)^\top \mathbf{W}\mathbf{p}^* - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W}\mathbf{p}^{\text{FPA}} \\
&\qquad + \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W}\mathbf{p}^{\text{FPA}} - \mathbf{q}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W}\mathbf{p}^{\text{FPA}} \\
&\le \left\|\left(\mathbf{q}^*(\mathbf{p}^*) - \boldsymbol{\mu}^*(\mathbf{p}^*)\right)^\top \mathbf{W}\mathbf{p}^*\right\|_1 + \left(\boldsymbol{\mu}^*(\mathbf{p}^*)^\top \mathbf{W}\mathbf{p}^* - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W}\mathbf{p}^{\text{FPA}}\right) \\
&\qquad + \left\|\left(\mathbf{q}^*(\mathbf{p}^{\text{FPA}}) - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})\right)^\top \mathbf{W}\mathbf{p}^{\text{FPA}}\right\|_1 \\
&\le \left\|\left(\mathbf{q}^*(\mathbf{p}^*) - \boldsymbol{\mu}^*(\mathbf{p}^*)\right)^\top \mathbf{W}\mathbf{p}^*\right\|_1 + \left\|\left(\mathbf{q}^*(\mathbf{p}^{\text{FPA}}) - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})\right)^\top \mathbf{W}\mathbf{p}^{\text{FPA}}\right\|_1 \\
&\le \left\|\mathbf{q}^*(\mathbf{p}^*) - \boldsymbol{\mu}^*(\mathbf{p}^*)\right\|_1 \left\|\mathbf{W}\mathbf{p}^*\right\|_\infty + \left\|\mathbf{q}^*(\mathbf{p}^{\text{FPA}}) - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})\right\|_1 \left\|\mathbf{W}\mathbf{p}^{\text{FPA}}\right\|_\infty \\
&\le 2C_\rho \max\left\{\left\|\mathbf{p}^*\right\|_\infty, \left\|\mathbf{p}^{\text{FPA}}\right\|_\infty\right\} \sqrt{n\|\mathcal{C}(G;\rho)\|_1},
\end{aligned}$$

where the first inequality follows trivially, the second inequality follows since the optimality of $\mathbf{p}^{\text{FPA}}$ for approximate pricing problem (22), the third inequality follows from Corollary 2, and the last one follows the properties of matrix operator norms. □

We then focus on the proof of the pricing problem in the adoption probability space.

Proof of Theorem 5: Let $\pi(\boldsymbol{\mu}) = \sum_{i\in V}\left(v_i + \beta\sum_{j\in\mathcal{N}_i}\frac{\mu_j}{d_i} + \ln\frac{1-\mu_i}{\mu_i}\right)\mu_i$. The Hessian matrix of $\pi(\boldsymbol{\mu})$ can be derived as

$$\frac{\partial^2\pi}{\partial\mu_i^2} = -\frac{1}{\gamma}\frac{1}{\mu_i(\mu_i-1)^2} \quad \text{and} \quad \frac{\partial^2\pi}{\partial\mu_i\partial\mu_j} = \frac{1}{\gamma}\mathbb{1}\{j\in\mathcal{N}_i\}\frac{\beta}{d_i} + \frac{1}{\gamma}\mathbb{1}\{i\in\mathcal{N}_j\}\frac{\beta}{d_j}.$$

For the diagonal elements of the Hessian matrix H_π, we can have $-1/[\mu(\mu-1)^2] \le -6.75$ holds for any $x\in[0,1]$. The inequality is tight when $\mu = 1/3$. For the nondiagonal elements of the Hessian matrix $\mathbf{H}_\pi$, we can find them related to the structure of network $G(V,E)$.

Therefore, we can have the Hessian matrix to be

$$\mathbf{H}_\pi = \frac{1}{\gamma}\text{diag}\left(\left\{-\frac{1}{\mu_i(\mu_i-1)^2}\right\}_{i\in V}\right) + \frac{\beta}{\gamma}\left(\tilde{\mathbf{A}}+\tilde{\mathbf{A}}^\top\right) \preceq \frac{1}{\gamma}\left[-6.75I+\beta\left(\tilde{\mathbf{A}}+\tilde{\mathbf{A}}^\top\right)\right]$$

By Gershgorin circle theorem, we can bound the eigenvalues of $\tilde{\mathbf{A}}$ by $-1 \le \lambda(\tilde{\mathbf{A}}) \le 1$. Since 1 is one of the eigenvalues of $\tilde{\mathbf{A}}$, we can have $\lambda_{\max}(\tilde{\mathbf{A}}) = 1$. Therefore, when $\beta \le 3.375$,

$$\lambda_{\max}\left(-6.75I+\beta\left(\tilde{\mathbf{A}}+\tilde{\mathbf{A}}^\top\right)\right) \le 0,$$

which implies that $\mathbf{H}_\pi$ is negative semi-definite. The equality holds when $\beta = 3.375$. □

The proof for the concavity of dynamic pricing objective largely mirrors that of the static pricing problem.

Proof of Theorem 6: Let $\pi(\boldsymbol{\mu}) = \sum_{i\in V}\left(v_i + \beta\sum_{j\in\mathcal{N}_i}\frac{\mu_j(t-1)}{d_i} + \ln\frac{1-\mu_i(t)}{\mu_i(t)}\right)\mu_i(t)$. The Hessian matrix of $\pi(\boldsymbol{\mu})$ can be derived as

$$\frac{\partial^2\pi}{\partial\mu_i(t)^2} = -\frac{1}{\gamma}\frac{1}{\mu_i(t)(\mu_i(t)-1)^2},\ \frac{\partial^2\pi}{\partial\mu_i(t)\partial\mu_j(t-1)} = \frac{1}{\gamma}\mathbb{1}\{j\in\mathcal{N}_i\}\frac{\beta}{d_i} \text{ and } \frac{\partial^2\pi}{\partial\mu_i(t-1)\partial\mu_j(t)} = \frac{1}{\gamma}\mathbb{1}\{i\in\mathcal{N}_j\}\frac{\beta}{d_j}$$

For the diagonal elements of the Hessian matrix H_π, we can have $-1/[\mu(\mu-1)^2] \le -6.75$ holds for any $x\in[0,1]$. The inequality is tight when $\mu = 1/3$. For the nondiagonal elements of the Hessian matrix $\mathbf{H}_\pi$, we can find them related to the network structure. Let us define the following block matrix:

$$\mathbf{M} := \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \tilde{\mathbf{A}} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{A}} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & & & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \tilde{\mathbf{A}} & \mathbf{0} \end{pmatrix}$$

Therefore, we can represent the Hessian matrix as the following block matrix

$$\frac{1}{\gamma}\begin{pmatrix} \text{diag}\left(\left\{-\frac{1}{\mu_i(\mu_i-1)^2}\right\}_{i\in V}\right) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \text{diag}\left(\left\{-\frac{1}{\mu_i(\mu_i-1)^2}\right\}_{i\in V}\right) & \cdots & \mathbf{0} \\ \vdots & & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \text{diag}\left(\left\{-\frac{1}{\mu_i(\mu_i-1)^2}\right\}_{i\in V}\right) \end{pmatrix} + \frac{\beta}{\gamma}\left(\mathbf{M}+\mathbf{M}^\top\right)$$

$$\preceq \frac{1}{\gamma}\left[-6.75\mathbf{I}+\beta\left(\mathbf{M}+\mathbf{M}^\top\right)\right].$$

By Gershgorin circle theorem, we can bound the eigenvalues of $\mathbf{M}$ by $-1 \le \lambda(\mathbf{M}) \le 1$. Since 1 is one of the eigenvalues of $\mathbf{M}$, we can have $\lambda_{\max}(\mathbf{M}) = 1$. Therefore, when $\beta \le 3.375$,

$$\lambda_{\max}\left(-6.75I+\beta\left(\mathbf{M}+\mathbf{M}^\top\right)\right) \le 0$$

which implies that $\mathbf{H}_\pi$ is negative semi-definite. The equality holds when $\beta = 3.375$. □

□

In the following, we illustrate the procedure of gradient descent for the pricing problem in the price space.

Gradient descent and approximate gradient descent. By taking the derivative on both sides of the fixed-point equation, we get

$$\frac{d\boldsymbol{\mu}(\mathbf{p})}{d\mathbf{p}} = \frac{\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))}{\partial\mathbf{p}} + \frac{d\boldsymbol{\mu}(\mathbf{p})}{d\mathbf{p}}\cdot\frac{\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))}{\partial\boldsymbol{\mu}(\mathbf{p})}.$$

By rearranging the terms, we obtain

$$\frac{d\boldsymbol{\mu}(\mathbf{p})}{d\mathbf{p}} \cdot \left(\mathbf{I} - \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})} \right) = \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \mathbf{p}}.$$

Matrix $(\mathbf{I} - \partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))/\partial \boldsymbol{\mu}(\mathbf{p}))$ is guaranteed to be invertible. The reason is that, by Proposition 2, we know that $\mathbf{h}$ is a contraction mapping and $\|\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))/\partial \boldsymbol{\mu}(\mathbf{p})\|_\infty < 1$.

With an eye toward implementation, we also notice that (26) requires computing the inverse of an n-by-n matrix. When the network is large and dense, this calculation becomes intimidating. Fortunately, we know that $\|\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))/\partial \boldsymbol{\mu}(\mathbf{p})\|_\infty < 1$, and hence, the spectral radius of $\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))/\partial \boldsymbol{\mu}(\mathbf{p})$ is smaller than 1. Consequently, we can expand the inverse as the Neumann series, which is similar to the centrality measure,

$$\left(\mathbf{I} - \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})} \right)^{-1} = \mathbf{I} + \sum_{\ell=1}^{\infty} \left(\frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})} \right)^{\ell}.$$

This leads to the following k-th order approximate gradient:

$$\frac{d\Pi(\mathbf{p})}{d\mathbf{p}} \approx \widetilde{\mathbf{G}}_k(\mathbf{p}) = \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \mathbf{p}} \cdot \left(\mathbf{I} + \sum_{\ell=1}^{k} \left(\frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})} \right)^{\ell} \right) \cdot \mathbf{W} \cdot \mathbf{p} + \mathbf{W}^\top \cdot \boldsymbol{\mu}(\mathbf{p}) \tag{44}$$

for the pricing problem. We expect such an easy-to-compute approximate gradient to lead to a significant efficiency gain, as is usually the case in the literature regarding approximate gradient descent (Ruder 2016). Previous works have applied similar low-order approximations for network effects for different purposes (e.g., see Candogan et al. 2012, Zeng et al. 2023). In subsequent numerical experiments, we find that $k = 2$ works very well in practice, leading to near-optimal solutions very quickly.

F.4. Experiments on Pricing Problems

In the experiments for pricing problems on a social network, the first issue is that the optimal pricing problem under the original diffusion model seems impossible to derive. We test over different randomly generated instances and find that the profits calculated via simulation and the FPA scheme are quite close, with a percentage error almost uniformly bounded by 0.5% in our experiment.

In order to check FPA performance about the total profit when price is considered, we test over three groups of instances. By fixing the expected number of neighbors to be 10, we generate diffusion instances with random graphs $G(20, 0.5)$, $G(100, 0.1)$, $G(1000, 0.01)$. For each instance, the agent is associated with an intrinsic value i.i.d. sampled from $\mathcal{U}(0, 4)$ and an offered price i.i.d. sampled from $\mathcal{U}(0, 4)$. We set $\beta = 3$ and $\gamma = 1$. In Figure 19, we show the distribution of profit difference among all diffusion instances. We notice that the absolute profit difference is small. Furthermore, as the network becomes larger, the performance gap becomes more concentrated. In conclusion, we can consider the FPA scheme to be able to achieve almost the same performance as the simulation methods in the pricing problem.

Hereafter, we compare the pricing scheme under the FPA scheme as default. We assume $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim}$ Logistic$(0, 1)$, which follows the theoretical analysis in Section 6.2.

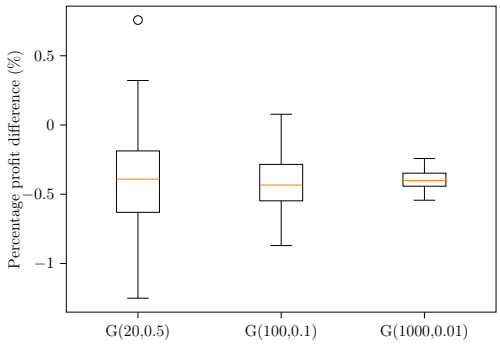


Figure 19 Profit difference between the ABS result and the FPA solution.

F.4.1. Static pricing problem. In this section, we consider the static pricing problem on a social network, considering limiting user behavior. We study two extreme scenarios, the *perfect price discrimination* case, where each consumer is offered a personal price, and the *public price* case, where all consumers receive the same price.

In the perfect price discrimination case, we test three different algorithms. The first algorithm is the gradient descent method in the adoption probability space (grad-PROB). With a network effect intensity that satisfies Theorem 5, grad-PROB can find the global optimal solution. The second algorithm is the gradient method in the price space (grad-PRICE). The third algorithm considers the pricing problem without network diffusion, that is, the price is determined according to the standard logit model. We still refer to it as the model misspecification (MM) scheme.

For the public price case, we also test three different algorithms. However, in this case, the pricing problem cannot be considered in the adoption probability space. Instead, we use a grid search (GS) to find an upper-bound solution to the problem. Specifically, we divide the price into grids of tolerance ξ. For each price p, we upper bound the profit with $(p-\xi)\cdot\sum_{i\in V}\mu_i(p)$. The other two algorithms, grad-PRICE and MM, as discussed above, are applied here.

For both scenarios, we test on a real-world network—*Amherst41* as introduced in Section 5.3. For each diffusion instance, we set the price sensitivity as $\gamma=0.1$, and the intrinsic value $v_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0,4)$. In Figure 20, we plot the realized profit of three algorithms against different values of the network effect intensity β. Furthermore, as remarked before, algorithm grad-PRICE involves the derivation of the gradient as (26), which requires calculating the inverse of a $n\times n$ matrix. We resort to the second-order approximate gradient, $\widetilde{G}_2(\mathbf{p})$, as given in (44).

We offer several observations from this figure. First, grad-PRICE obtains a near-optimal solution in the case of price discrimination. This hints that we can use grad-PRICE to gain high-quality results in the general pricing setting when grad-PROB is not applicable. Second, there is a significant performance degradation of MM when the network effect is large. When $\beta=3$, the relative profit loss reaches 21.16% and 8.30% if the network effect is ignored, respectively. When β increases to over 3.375 that violates the condition in 5, we can see that both algorithms grad-PROB and grad-PRICE still perform exceptionally. Third, comparing these two cases, we find that pricing discrimination can significantly increase the total profit, especially when the network effect intensity is large.

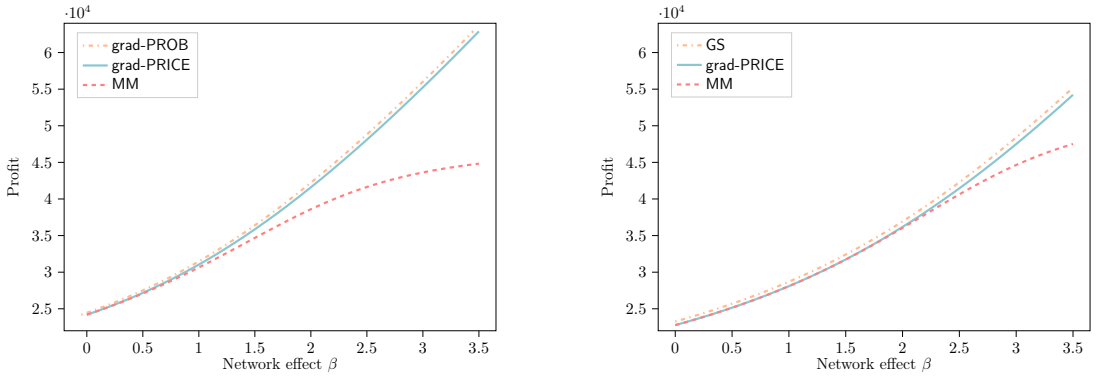


Figure 20 Realized profit versus network effect intensity. Left: Perfect price discrimination; Right: Public price (The curve of grad-PROB coincides with grad-PRICE in the left subfigure. In order to make them identifiable in the figures, we shift the grad-PROB to the left by 0.05.)

Furthermore, we compare the performance of these algorithms in more instances in terms of their execution time and the quality of solutions. We assume parameters $(n, p(n), \beta) \in \{100, 1000\} \times \{\frac{1}{n}, \frac{\log n}{n}, 0.1\} \times \{2, 3.5\}$. The numerical results for the two cases are shown in Tables 5 and 6. The grad-PRICE approach derives high-quality solutions in both cases. We notice that in the perfect price discrimination case, the run time for grad-PROB is less than grad-PRICE, although the margin is not too large. The profit difference between these two approaches is quite small, uniformly smaller than 0.25%. For the public price case, we set the tolerance of the grid search to be 0.5 within the range $[0, 100]$. grad-PRICE runs much faster than the grid search with a performance loss of up to 2%. The performance of MM remains poor across the two cases in this experiment, suggesting that the loss from ignoring network effects can be detrimental. Moreover, in the case when $\beta = 3.5$, we notice that the condition in 5 is violated, but practically, our algorithms provide commendable results. In summary, our main message through the numerical experiments is twofold. First, it is important to incorporate the network effect into operational problems. The gain from doing so can be significant. Second, we advocate grad-PRICE as a practical method for price optimization. With our approximate gradient expression tailored to the network setting as in (44), grad-PRICE becomes a competitive price optimization technique. It can be efficiently implemented in various practical scenarios to find high-quality price solutions.

F.4.2. Dynamic pricing problem .

In this section, we investigate the outcomes of applying dynamic pricing, as opposed to deriving the static optimal price once equilibrium is reached, under the perfect price discrimination case.

We evaluate three distinct pricing schemes: static, dynamic, and myopic. The static scheme utilizes the optimal static price derived in Section 6.2.1, assuming limiting consumer behavior. In contrast, the Dynamic scheme incorporates transient consumer behavior over time as in Section 6.2.2. Additionally, we introduce the Myopic benchmark, where the optimal price is recalculated at each time step based on the adoption levels

Table 5 Numerical results of pricing problem for randomly generated instances (perfect price discrimination)

Parameters $(n, p(n), \beta)$	grad-PROB	grad-PRICE				MM			
	time (s)	time (s)	profit loss (%)			time (s)	profit loss (%)		
			min	mean	max		min	mean	max
(100,1/n,2)	0.067	0.160	0.007	0.025	0.169	0.002	5.887	7.118	8.070
(100,log n/n,2)	0.056	0.154	0.026	0.038	0.053	0.002	5.717	6.670	7.759
(100,0.1,2)	0.044	0.128	0.031	0.040	0.058	0.002	5.238	6.199	6.990
(100,1/n,3.5)	0.034	0.090	0.247	0.023	0.842	0.001	19.824	22.625	25.161
(100,log n/n,3.5)	0.072	0.334	0.061	0.084	0.177	0.002	27.406	28.160	28.807
(100,0.1,3.5)	0.035	0.090	0.039	0.197	0.802	0.001	19.550	22.800	24.965
(1,000,1/n,2)	0.785	2.422	0.010	0.031	0.095	0.100	6.769	7.037	7.358
(1,000,log n/n,2)	0.576	1.986	0.035	0.039	0.043	0.099	6.139	6.390	6.697
(1,000,0.1,2)	0.696	1.917	0.037	0.041	0.047	0.099	5.376	5.878	6.251
(1,000,1/n,3.5)	0.616	1.658	0.114	0.233	0.429	0.080	21.794	22.798	23.591
(1,000,log n/n,3.5)	0.673	4.648	0.079	0.086	0.095	0.104	27.747	28.050	28.377
(1,000,0.1,3.5)	0.107	1.657	0.088	0.097	0.110	0.075	27.273	27.666	28.071

Table 6 Numerical results of pricing problem for randomly generated instances (public price)

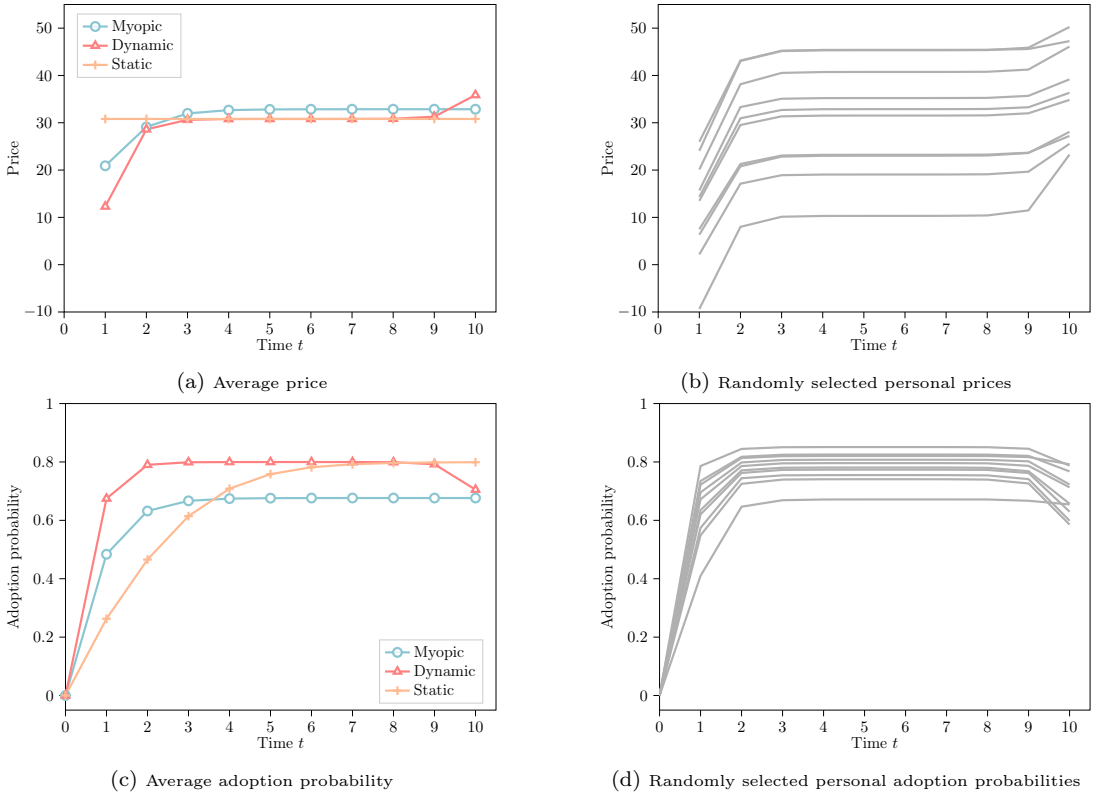
Parameters $(n, p(n), \beta)$	GS	grad-PRICE				MM			
	time (s)	time (s)	profit loss (%)			time (s)	profit loss (%)		
			min	mean	max		min	mean	max
(100,1/n,2)	0.893	0.084	1.882	2.050	2.276	0.001	2.048	2.203	2.539
(100,log n/n,2)	1.364	0.087	1.882	2.226	0.013	0.001	2.085	2.411	3.051
(100,0.1,2)	1.075	0.073	1.874	2.038	2.260	0.001	2.088	2.441	3.036
(100,1/n,3.5)	1.065	0.083	1.721	1.870	2.029	0.001	2.608	4.216	6.380
(100,log n/n,3.5)	1.866	0.130	1.577	1.766	2.088	0.001	9.361	13.821	18.284
(100,0.1,3.5)	1.454	0.111	1.558	1.725	1.944	0.001	11.177	14.465	18.720
(1,000,1/n,2)	10.376	0.942	1.991	2.058	2.018	0.002	2.072	2.145	2.245
(1,000,log n/n,2)	12.891	0.916	1.981	2.036	2.100	0.002	2.235	2.402	2.587
(1,000,0.1,2)	13.057	0.975	1.971	2.027	2.093	0.002	2.247	2.408	2.594
(1,000,1/n,3.5)	12.144	0.902	1.791	1.864	1.942	0.002	3.285	4.192	4.983
(1,000,log n/n,3.5)	14.360	1.389	1.710	1.764	1.839	0.002	12.175	13.806	15.060
(1,000,0.1,3.5)	15.053	1.442	1.681	1.736	1.819	0.002	12.816	14.145	15.525

from the previous period. Specifically, given $\boldsymbol{\mu}(t-1)$, we solve the following optimization problem at each time step t:

$$\begin{aligned} \underset{\boldsymbol{\mu}(t)}{\text{maximize}} \quad & \sum_{i\in V} \frac{1}{\gamma}\left(v_i + \beta \sum_{j\in\mathcal{N}_i} \frac{\mu_j(t-1)}{d_i} + \ln\frac{1-\mu_i(t)}{\mu_i(t)}\right)\mu_i(t) \\ \text{subject to} \quad & 0 \le \mu_i(t) \le 1,\ \forall i \in V. \end{aligned}$$

We still use the real-world network—*Amherst41* as the numerical illustration, similar to the experiment in Appendix F.4. We set the price sensitivity to $\gamma = 0.1$, the intrinsic value $v_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0,4)$, and the network effect intensity $\beta = 3$ in order to satisfy the condition in Theorem 6. In Figures 21a and 21c, we present the averages of optimal dynamic prices and adoption probabilities across all agents, respectively. The results show that the optimal pricing strategy involves using price as a leverage tool to quickly drive agents' behavior toward the optimal static case within the first few time steps (approximately 3 steps), after which the price stabilizes at the optimal static level. Despite the negative impact that price should typically have on adoption, the

existence of network effect causes both price and adoption probability to rise in the early stages under the optimal pricing strategy. Similarly, the myopic pricing scheme also sees an initial increase in both price and adoption, but its stabilized price is slightly higher than the optimal static price, resulting in suboptimal long-term behavior. In Figures 21b and 21d, we display the price and adoption probability trajectories for 10 randomly selected agents from a specific network instance. Under perfect price discrimination, different agents experience distinct prices, but their price trajectories follow a pattern similar to the average price. This dynamic pricing pattern suggests that, with network effects, the platform first uses a low price to attract more customers and then stabilizes at the optimal static price to maximize long-term profit.



(a) Average price

(b) Randomly selected personal prices

(c) Average adoption probability

(d) Randomly selected personal adoption probabilities

Figure 21 Different dynamic pricing schemes under the perfect price discrimination.

Finally, in Figure 22, we demonstrate the performance of different pricing schemes. Both the myopic and dynamic schemes yield higher profits than the static scheme during the early time steps (i.e., the transient phase). However, as time progresses, these gains diminish. In the long run, the dynamic scheme converges to the same performance as the static scheme, while the myopic scheme remains slightly lower.

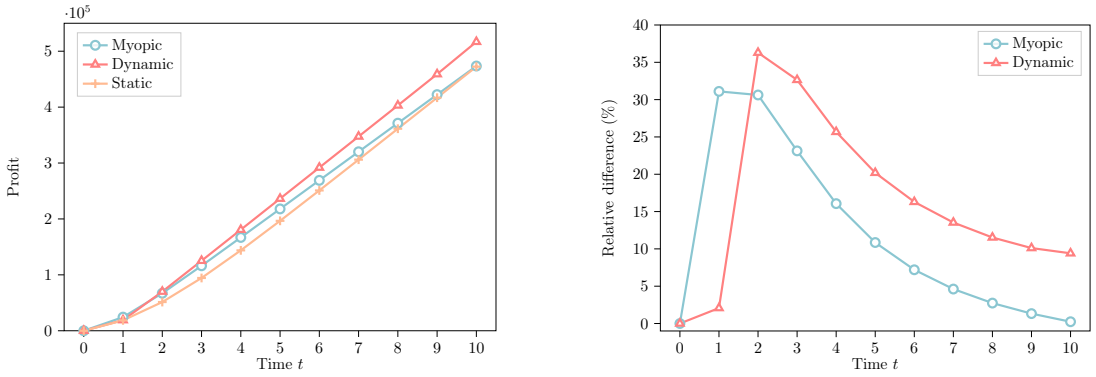


Figure 22 **The performance of different dynamic pricing schemes under the perfect price discrimination. Left: Accumulative profit; Right: Relative difference in profit where Static is set as the benchmark.**